

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:25:44 ; Search time 55.8955 Seconds
 (without alignments)
 540.877 Million cell updates/sec

Title: US-09-423-100-6
 Perfect score: 587
 Sequence: 1 MFPTIPLSRLFDNAMLRAHR.....IVEQCCTSICSILYQLENYCN 107

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 1586107 seqs, 282547505 residues

Total number of hits satisfying chosen parameters: 1586107

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : A_Geneseq_29Jan04:*
 1: geneseqp1980s:*
 2: geneseqp1990s:*
 3: geneseqp2000s:*
 4: geneseqp2001s:*
 5: geneseqp2002s:*
 6: geneseqp2003as:*
 7: geneseqp2003bs:*
 8: geneseqp2004s:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query		DB	ID	Description
		Match	Length			
1	587	100.0	107	2	AA42860	Aay42860 hGH-mini-
2	555.5	94.6	150	2	AA42861	Aay42861 Chimeric
3	315.5	53.7	116	2	AAR98897	Aar98897 SOD-proin
4	304	51.8	63	2	AAR68900	Aar68900 Human pro
5	304	51.8	117	2	AAR98896	Aar98896 SOD-proin
6	302.5	51.5	137	2	AAR71692	Aar71692 Mating fa
7	299	50.9	56	2	AAR68901	Aar68901 Human pro
8	299	50.9	56	2	AAR78665	Aar78665 Proinsuli
9	299	50.9	96	2	AAR68899	Aar68899 Human pro

10	299	50.9	96	2	AAR78662	Aar78662	Fusion pr
11	299	50.9	145	2	AAR71694	Aar71694	Mating fa
12	299	50.9	146	2	AAR71695	Aar71695	Mating fa
13	294	50.1	52	2	AAY42859	Aay42859	Human ins
14	293	49.9	57	2	AAR04582	Aar04582	Proinsuli
15	288.5	49.1	160	2	AAR79056	Aar79056	Glycosylp
16	287	48.9	52	2	AAR11899	Aar11899	Example o
17	287	48.9	65	2	AAW47365	Aaw47365	Preproins
18	287	48.9	138	2	AAR87086	Aar87086	pKV142 mo
19	284.5	48.5	58	2	AAR96047	Aar96047	Modified
20	284.5	48.5	59	2	AAR96048	Aar96048	Modified
21	284.5	48.5	65	2	AAR88188	Aar88188	N-termina
22	284.5	48.5	109	1	AAP94645	Aap94645	Amino aci
23	284.5	48.5	123	2	AAW19240	Aaw19240	EEAEPK-MI
24	284.5	48.5	123	2	AAW69160	Aaw69160	DNA const
25	284.5	48.5	124	2	AAW78751	Aaw78751	pAK855 pr
26	284.5	48.5	124	6	ABP55059	Abp55059	Insulin p
27	284.5	48.5	124	6	ABB82578	Abb82578	Synthetic
28	284.5	48.5	125	2	AAW19242	Aaw19242	EEAEPK-MI
29	284	48.4	138	1	AAP94643	Aap94643	Amino aci
30	284	48.4	138	2	AAW04890	Aaw04890	S. cerevi
31	284	48.4	140	2	AAR71693	Aar71693	Mating fa
32	284	48.4	140	2	AAR71690	Aar71690	Mating fa
33	283.5	48.3	53	2	AAR65883	Aar65883	Di-Arg-(B
34	283.5	48.3	53	2	AAW18007	Aaw18007	Ins1 doub
35	283.5	48.3	117	2	AAW78752	Aaw78752	Protein s
36	283.5	48.3	408	4	AAB30705	Aab30705	A Bacillu
37	282	48.0	94	1	AAP94644	Aap94644	Amino aci
38	282	48.0	120	2	AAW19241	Aaw19241	EEAEPK-MI
39	281.5	48.0	60	1	AAP20002	Aap20002	Human pro
40	281.5	48.0	105	1	AAP94648	Aap94648	Amino aci
41	281.5	48.0	153	3	AAY53589	Aay53589	Human pre
42	281	47.9	102	1	AAP94649	Aap94649	Signal-le
43	281	47.9	104	2	AAR71684	Aar71684	Yeast sig
44	280.5	47.8	55	1	AAP71019	Aap71019	Sequence
45	280.5	47.8	89	2	AAR88179	Aar88179	Signal pe

ALIGNMENTS

RESULT 1

AAY42860

ID AAY42860 standard; protein; 107 AA.

XX

AC AAY42860;

XX

DT 19-JAN-2000 (first entry)

XX

DE hGH-mini-proinsulin chimeric protein.

XX

KW Insulin; precursor; growth hormone; chaperone; intramolecular; folding;

KW conformation; chimeric protein; cleavable; recombinant; production;

KW yield.

XX

OS Synthetic.

OS Homo sapiens.

XX
 PN WO9950302-A1.
 XX
 PD 07-OCT-1999.
 XX
 PF 31-MAR-1998; 98WO-CN000052.
 XX
 PR 31-MAR-1998; 98WO-CN000052.
 XX
 PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.
 XX
 PI Gan Z;
 XX
 DR WPI; 1999-610839/52.
 XX
 PT New chimeric proteins containing human growth hormone fragment, used
 PT particularly for the production of human insulin.
 XX
 PS Claim 13; Page 30; 46pp; English.
 XX
 CC This sequence represents a chimeric protein, hGH-mini-proinsulin. This
 CC chimeric protein contains an N-terminal fragment of human growth hormone
 CC (hGH) of the sequence given in AAY42855, a cleavable peptide linker
 CC (AAY42857), and a human insulin precursor comprising insulin A and B
 CC chains (AAY42859). The hGH portion of the chimeric protein acts as an
 CC intramolecular chaperone (IMC) for the insulin precursor, enabling it to
 CC fold correctly. The cleavable peptide linker has a C-terminal Arg residue
 CC which enables the hGH portion of the chimeric protein to be removed after
 CC folding has taken place. Production of recombinant human insulin via an
 CC hGH-proinsulin chimeric protein can provide human insulin with correctly
 CC linked cysteine bridges with fewer necessary procedural steps, and hence
 CC resulting in a higher yield of human insulin. The IMC sequences not only
 CC protect insulin sequences from intracellular degradation by a
 CC microorganism host, but also promote the folding of the fused insulin
 CC precursor, facilitate the solubility of the fusion protein and decrease
 CC the intermolecular interactions among the fusion proteins, thus allowing
 CC folding of the fused insulin precursor at commercially useful high
 CC concentrations. The procedural steps of cyanogen bromide cleavage,
 CC oxidative sulphytolysis and related purification steps can thus be
 CC eliminated, along with the use of high concentrations of mercaptan or the
 CC use of hydrophobic absorbent resins
 XX
 SQ Sequence 107 AA;

Query Match 100.0%; Score 587; DB 2; Length 107;
 Best Local Similarity 100.0%; Pred. No. 2.4e-43;
 Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPLGTGPRFVNQH 60
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 1 MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPLGTGPRFVNQH 60

 QY 61 LCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 61 LCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107

RESULT 2

AAY42861

ID AAY42861 standard; protein; 150 AA.

XX

AC AAY42861;

XX

DT 19-JAN-2000 (first entry)

XX

DE Chimeric protein, SEQ ID 7.

XX

KW Insulin; precursor; growth hormone; chaperone; intramolecular; folding;
KW conformation; chimeric protein; cleavable; recombinant; production;
KW yield.

XX

OS Synthetic.

OS Homo sapiens.

XX

PN WO9950302-A1.

XX

PD 07-OCT-1999.

XX

PF 31-MAR-1998; 98WO-CN000052.

XX

PR 31-MAR-1998; 98WO-CN000052.

XX

PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX

PI Gan Z;

XX

DR WPI; 1999-610839/52.

XX

PT New chimeric proteins containing human growth hormone fragment, used
PT particularly for the production of human insulin.

XX

PS Claim 14; Page 30-31; 46pp; English.

XX

CC This sequence represents a chimeric protein, which contains an N-terminal
CC fragment of human growth hormone (hGH) of the sequence given in AAY42856,
CC a cleavable peptide linker (AAY42857), and a human insulin precursor
CC comprising insulin A and B chains (AAY42859). The hGH portion of the
CC chimeric protein acts as an intramolecular chaperone (IMC) for the
CC insulin precursor, enabling it to fold correctly. The cleavable peptide
CC linker has a C-terminal Arg residue which enables the hGH portion of the
CC chimeric protein to be removed after folding has taken place. Production
CC of recombinant human insulin via an hGH-proinsulin chimeric protein can
CC provide human insulin with correctly linked cysteine bridges with fewer
CC necessary procedural steps, and hence resulting in a higher yield of
CC human insulin. The IMC sequences not only protect insulin sequences from
CC intracellular degradation by a microorganism host, but also promote the
CC folding of the fused insulin precursor, facilitate the solubility of the
CC fusion protein and decrease the intermolecular interactions among the
CC fusion proteins, thus allowing folding of the fused insulin precursor at
CC commercially useful high concentrations. The procedural steps of cyanogen
CC bromide cleavage, oxidative sulphytolysis and related purification steps
CC can thus be eliminated, along with the use of high concentrations of
CC mercaptan or the use of hydrophobic absorbent resins

XX

SQ Sequence 150 AA;

Query Match 94.6%; Score 555.5; DB 2; Length 150;
Best Local Similarity 71.3%; Pred. No. 1.7e-40;
Matches 107; Conservative 0; Mismatches 0; Indels 43; Gaps 1;

```
Qy      1 MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP----- 49
          ||||||||||||||||||||||||||||||||||||||||||||
Db      1 MFPTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNPQTSLSFSSESIP 60

Qy     50 -----LGTGPRFVNQHLCGSHLVEALYLVCGER 77
          ||||||||||||||||||||||||||||||||||||||||
Db     61 TPSNREETQOKSNLELLLRISLLLIQSWLEPVQLGTGPRFVNQHLCGSHLVEALYLVCGER 120

Qy     78 GFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
          ||||||||||||||||||||||||||||
Db    121 GFFYTPKTRGIVEQCCTSICSLYQLENYCN 150
```

RESULT 3

AAR98897

ID AAR98897 standard; protein; 116 AA.

XX

AC AAR98897;

XX

DT 03-FEB-1997 (first entry)

XX

DE SOD-proinsulin hybrid polypeptide.

XX

KW Insulin; proinsulin; hybrid polypeptide; protein folding;

KW enzymatic cleavage; cyanogen bromide; sulphitolysis.

XX

OS Homo sapiens..

XX

PN W09620724-A1.

XX

PD 11-JUL-1996.

XX

PF 29-DEC-1994; 94WO-US013268.

XX

PR 29-DEC-1994; 94WO-US013268.

XX

PA (BIOT-) BIO-TECHNOLOGY GENERAL CORP.

XX

PI Hartman JR, Mendelovitz S, Gorecki M;

XX

DR WPI; 1996-333766/33.

DR N-PSDB; AAT34670.

XX

PT Recombinant insulin prodn. by correctly folding pro-insulin hybrid

PT polypeptide - then enzymatic cleavage of folded product, does not require

PT sulphite protection of SH nor use of cyanogen bromide.

XX

PS Example 1B; Fig 7; 69pp; English.

XX

CC A new method for the production of recombinant human insulin comprises

CC folding a hybrid polypeptide comprising proinsulin under conditions that

CC permit correct disulphide bond formation and subjecting that folded
CC protein to enzymatic cleavage. The insulin produced can then be purified.
CC This sequence is a SOD-insulin B chain-Arg-insulin A chain hybrid
CC polypeptide and is encoded by the plasmid construct pDBAST-LAT.
CC Transformation of the proper E.coli host cells with pDBAST-LAT results in
CC the efficient expression of the proinsulin hybrid polypeptide, useful for
CC human insulin production. The method produces recombinant human insulin
CC identical to the natural hormone. Hazardous and cumbersome procedures
CC involving cyanogen bromide and sulphitolysis to protect SH groups are
CC avoided since the entire hybrid polypeptide folds efficiently to the
CC native structure even with the leader attached and Cys unprotected
XX
SQ Sequence 116 AA;

```

RESULT 4
AAR68900
ID    AAR68900 standard; peptide; 63 AA.
XX
AC    AAR68900;
XX
DT    25-MAR-2003   (revised)
DT    02-MAR-1995   (first entry)
XX
DE    Human pro-insulin 4.
XX
KW    Pro-insulin; A-chain; B-chain; C-chain; disulphide; mercaptan;
KW    chaotropic agent.
XX
OS    Homo sapiens.
XX
PN    EP600372-A1.
XX
PD    08-JUN-1994.
XX
PF    25-NOV-1993;   93EP-00118993.
XX
PR    02-DEC-1992;   92DE-04240420.
XX
PA    (FARH ) HOECHST AG.
XX
PI    Obermeier R,   Gerl M,   Ludwig J,   Sabel W;
XX
DR    WPI; 1994-177718/22.
XX

```


DR WPI; 1996-333766/33.
DR N-PSDB; AAT34669.
XX
PT Recombinant insulin prodn. by correctly folding pro-insulin hybrid
PT polypeptide - then enzymatic cleavage of folded product, does not require
PT sulphite protection of SH nor use of cyanogen bromide.
XX
PS Example 1A; Fig 6; 69pp; English.
XX
CC A new method for the production of recombinant human insulin comprises
CC folding a hybrid polypeptide comprising proinsulin under conditions that
CC permit correct disulphide bond formation and subjecting that folded
CC protein to enzymatic cleavage. The insulin produced can then be purified.
CC This sequence is a SOD-insulin B chain-Lys-Arg-insulin A chain hybrid
CC polypeptide and is encoded by the plasmid construct pBAST-R.
CC Transformation of the proper E.coli host cells with pBAST-R results in
CC the efficient expression of the proinsulin hybrid polypeptide, useful for
CC human insulin production. The method produces recombinant human insulin
CC identical to the natural hormone. Hazardous and cumbersome procedures
CC involving cyanogen bromide and sulphitolysis to protect SH groups are
CC avoided since the entire hybrid polypeptide folds efficiently to the
CC native structure even with the leader attached and Cys unprotected
XX
SQ Sequence 117 AA;

RESULT 6

ID AAR71692 standard; protein; 137 AA.

AC AAR71692;

DT 25-MAR-2003 (revised)

DT 20-NOV-1995 (first entry)

DE Mating factor alpha 1-Insulin precursor ArgB31.

KW Human insulin precursor ArgB31; diabetes; Zinc ion complex;

XX

XX

FT	Protein	1. .85
----	---------	--------

```
FT          /label= mating factor alpha-1
```


FT Peptide 86. .116
 FT /label= B-chain
 FT Peptide 117. .137
 FT /label= A-chain
 XX
 PN WO9507931-A1.
 XX
 PD 23-MAR-1995.
 XX
 PF 16-SEP-1994; 94WO-DK000347.
 XX
 PR 17-SEP-1993; 93DK-00001044.
 PR 02-FEB-1994; 94US-00190829.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Havelund S, Halstrom JB, Jonassen I, Andersen AS, Markussen J;
 XX
 DR WPI; 1995-131314/17.
 DR N-PSDB; AAQ86425.
 XX
 PT Acylated insulin deriv. which may be present as a Zinc ion complex - is
 PT used to treat diabetes and is rapid acting.
 XX
 PS Example 5; Page 78; 100pp; English.
 XX
 CC AAQ86425 encodes AAR71692 mating factor alpha 1-Insulin precursor ArgB31.
 CC ArgB31 comprises the B and A chains of a claimed human insulin
 CC derivative. In the final claimed compsn. they are covalently connected
 CC via disulphide bonds between Cys residues A7/B7 and A20/B19. The
 CC derivative, which may be present as a zinc ion complex, can be used as a
 CC fast action treatment for diabetes. (Updated on 25-MAR-2003 to correct PN
 CC field.)
 XX
 SQ Sequence 137 AA;

Query Match 51.5%; Score 302.5; DB 2; Length 137;
 Best Local Similarity 50.0%; Pred. No. 1.1e-18;
 Matches 70; Conservative 4; Mismatches 27; Indels 39; Gaps 4;

Qy 2 FPTIPLSRFLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQ--KYSFLQ-----N 48
 ||:| | : | :| ||||| | ||: |
 Db 3 FPSI-----FTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLFFSN 57

 Qy 49 PLGTG-----PRFVNQHLGSHLVEALYLVCGERGFFYTPKTRG 87
 | |||||
 Db 58 STNNGLLFINTTIASIAAKEEGVSMARFVNQHLGSHLVEALYLVCGERGFFYTPKTRG 117

 Qy 88 IVEQCCTSICSLYQLENYCN 107
 |||||
 Db 118 IVEQCCTSICSLYQLENYCN 137

RESULT 7
 AAR68901
 ID AAR68901 standard; peptide; 56 AA.
 XX

AC AAR68901;
 XX
 DT 25-MAR-2003 (revised)
 DT 02-MAR-1995 (first entry)
 XX
 DE Human pro-insulin 3.
 XX
 KW Pro-insulin; A-chain; B-chain; C-chain; disulphide; mercaptan;
 KW chaotropic agent.
 XX
 OS Homo sapiens.
 XX
 PN EP600372-A1.
 XX
 PD 08-JUN-1994.
 XX
 PF 25-NOV-1993; 93EP-00118993.
 XX
 PR 02-DEC-1992; 92DE-04240420.
 XX
 PA (FARH) HOECHST AG.
 XX
 PI Obermeier R, Gerl M, Ludwig J, Sabel W;
 XX
 DR WPI; 1994-177718/22.
 XX
 PT Prodn. of pro-insulin with correct di:sulphide bridges - by treating
 PT recombinant precursor protein with mercaptan in alkali and in presence of
 PT chaotropic agent, then isolation on hydrophobic resin.
 XX
 PS Disclosure; Page 12; 15pp; German.
 XX
 CC Pro-insulin is produced by treating recombinant precursor protein with a
 CC mercaptan to provide 2-10 SH residues per Cys residue, in presence of a
 CC chaotropic agent and in aq. medium of pH 10-11, treating the prod. with 3
 CC -50 g hydrophobic adsorber resin per l aq. medium of pH 4-7, isolating
 CC the adsorbed resin and pro-insulin and desorbing the pro-insulin. This
 CC method produces pro-insulin with correctly bonded Cys bridges. Compared
 CC with known methods it involves fewer stages (esp. no sulphitolysis or
 CC cyanogen bromide cleavage) and overall losses during purification are
 CC reduced, i.e. the process is quicker and gives better yields. Sequences
 CC of insulin chain A, B and C are given in AAR68895-97. Sequences of pro-
 CC insulin 1-4 are given in AAR68898-901. (Updated on 25-MAR-2003 to correct
 CC PN field.)
 XX
 SQ Sequence 56 AA;

Query Match 50.9%; Score 299; DB 2; Length 56;
 Best Local Similarity 100.0%; Pred. No. 9.5e-19;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 |||||
 Db 4 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 56

RESULT 8

AAR78665

ID AAR78665 standard; protein; 56 AA.

XX

AC AAR78665;

XX

DT 03-APR-1996 (first entry)

XX

DE Proinsulin sequence 3.

XX

KW Proinsulin; post-translational modification; recombinant production;
KW protein folding; conformation.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Region 1. .4

FT /label= R2

FT /note= "a peptide of 4 amino acids"

FT Peptide 5. .34

FT /label= R1-(B2-B29)-Y

FT /note= "human insulin B-chain"

FT Region 35

FT /label= X

FT Peptide 36. .56

FT /label= Gly-(A2-A20)-R3

FT /note= "human insulin A-chain"

XX

PN EP668292-A2.

XX

PD 23-AUG-1995.

XX

PF 09-FEB-1995; 95EP-00101748.

XX

PR 18-FEB-1994; 94DE-04405179.

XX

PA (FARH) HOECHST AG.

XX

PI Obermeier R, Gerl M, Ludwig J, Sabel W;

XX

DR WPI; 1995-284754/38.

XX

PT Isolation of insulin that is correctly post-translationally processed -
PT by reacting pro:insulin with a mercaptan in the presence of a chaotropic
PT agent and purificn. after absorption to hydrophobic resin.

XX

PS Example 2; Page 13; 16pp; German.

XX

CC The present sequence is an example of a proinsulin molecule corresp. to
CC the general formula R2-R1-(B2-B29)-Y-X-Gly-(A2-A20)-R3 (II). In formula
CC (II), X = Lys, Arg or a peptide of 2-35 amino acids contg. Lys or Arg at
CC the N- and C-termini; Y = a natural amino acid; R1 = Phe or a bond; R2 =
CC H, Arg, Lys, a peptide of 2-45 amino acids contg. Arg or Lys at the N-
CC and C-termini; R3 = a natural amino acid; (A2-A20) and (B2-B29) are the
CC insulin A- and B-chain sequences from human or other insulin. The
CC proinsulin molecule (produced in recombinant E.coli) is reacted with
CC mercaptan at a ratio of 2-10 SH residues of mercaptan per Cys residue of
CC proinsulin. The reaction takes place in the presence of a chaotropic

CC method produces pro-insulin with correctly bonded Cys bridges. Compared
CC with known methods it involves fewer stages (esp. no sulphitolysis or
CC cyanogen bromide cleavage) and overall losses during purification are
CC reduced, i.e. the process is quicker and gives better yields. Sequences
CC of insulin chain A, B and C are given in AAR68895-97. Sequences of pro-
CC insulin 1-4 are given in AAR68898-901. (Updated on 25-MAR-2003 to correct
CC PN field.)

XX

SQ Sequence 96 AA;

Query Match 50.9%; Score 299; DB 2; Length 96;
Best Local Similarity 100.0%; Pred. No. 1.5e-18;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|
Db 44 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 96

RESULT 10

AAR78662

ID AAR78662 standard; protein; 96 AA.

XX

AC AAR78662;

XX

DT 03-APR-1996 (first entry)

XX

DE Fusion protein contg. proinsulin sequence 3.

XX

KW Proinsulin; post-translational modification; recombinant production;
KW protein folding; conformation.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Region 41. .44

FT /label= R2

FT /note= "a peptide of 4 amino acids"

FT Peptide 45. .74

FT /label= R1-(B2-B29)-Y

FT /note= "human insulin B-chain"

FT Region 75

FT /label= X

FT Peptide 76. .96

FT /label= Gly-(A2-A20)-R3

FT /note= "human insulin A-chain"

XX

PN EP668292-A2.

XX

PD 23-AUG-1995.

XX

PF 09-FEB-1995; 95EP-00101748.

XX

PR 18-FEB-1994; 94DE-04405179.

XX

PA (FARH) HOECHST AG.

XX

PI Obermeier R, Gerl M, Ludwig J, Sabel W;
 XX
 DR WPI; 1995-284754/38.
 XX
 PT Isolation of insulin that is correctly post-translationally processed -
 PT by reacting pro:insulin with a mercaptan in the presence of a chaotropic
 PT agent and purificn. after absorbtion to hydrophobic resin.
 XX
 PS Example 2; Page 8; 16pp; German.
 XX
 CC The present sequence is that of a fusion protein, produced in E.coli
 CC which contains an example of a proinsulin molecule corresp. to the
 CC general formula R2-R1-(B2-B29)-Y-X-Gly-(A2-A20)-R3 (II). In formula (II),
 CC X = Lys, Arg or a peptide of 2-35 amino acids contg. Lys or Arg at the N-
 CC and C-termini; Y = a natural amino acid; R1 = Phe or a bond; R2 = H, Arg,
 CC Lys, a peptide of 2-45 amino acids contg. Arg or Lys at the N- and C-
 CC termini; R3 = a natural amino acid; (A2-A20) and (B2-B29) are the insulin
 CC A- and B-chain sequences from human or other insulin. The proinsulin
 CC molecule, released by cyanogen bromide, is reacted with mercaptan at a
 CC ratio of 2-10 SH residues of mercaptan per Cys residue of proinsulin. The
 CC reaction takes place in the presence of a chaotropic auxiliary agent at
 CC pH 10-11 and results in proinsulin with correctly linked cystine bridges.
 CC Reaction with trypsin and opt. carboxypeptidase B yields correctly folded
 CC insulin. The insulin is isolated by absorbtion on a hydrophobic resin
 XX
 SQ Sequence 96 AA;

Query Match 50.9%; Score 299; DB 2; Length 96;
 Best Local Similarity 100.0%; Pred. No. 1.5e-18;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 44 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 96

RESULT 11

AAR71694

ID AAR71694 standard; protein; 145 AA.

XX

AC AAR71694;

XX

DT 25-MAR-2003 (revised)

DT 20-NOV-1995 (first entry)

XX

DE Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.

XX

KW Human insulin precursor ArgB1, ArgB31; diabetes; Zinc ion complex;

KW mating factor alpha 1; N-terminal EEAEAEAR.

XX

OS Homo sapiens.

XX

FH Key Location/Qualifiers

FT Protein 1. .85

FT /label= mating factor alpha-1

FT Peptide 86. .93

FT /label= N-terminal peptide

FT Peptide 94. .124
 FT /label= B-chain
 FT Peptide 125. .145
 FT /label= A-chain
 XX
 PN WO9507931-A1.
 XX
 PD 23-MAR-1995.
 XX
 PF 16-SEP-1994; 94WO-DK000347.
 XX
 PR 17-SEP-1993; 93DK-00001044.
 PR 02-FEB-1994; 94US-00190829.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Havelund S, Halstrom JB, Jonassen I, Andersen AS, Markussen J;
 XX
 DR WPI; 1995-131314/17.
 DR N-PSDB; AAQ86429.
 XX
 PT Acylated insulin deriv. which may be present as a Zinc ion complex - is
 PT used to treat diabetes and is rapid acting.
 XX
 PS Example 5; Page 82-83; 100pp; English.
 XX
 CC AAQ86429 encodes AAR71694 mating factor alpha 1-Insulin precursor ArgB1,
 CC ArgB31 N-terminal EEAEAEAR. The insulin precursor comprises the B and A
 CC chains of a claimed human insulin derivative preceded by the N-terminal
 CC amino acids EEAEAEAR. In the final claimed compsn. they are covalently
 CC connected via disulphide bonds between Cys residues A7/B7 and A20/B19.
 CC The derivative, which may be present as a zinc ion complex, can be used
 CC as a fast action treatment for diabetes. (Updated on 25-MAR-2003 to
 CC correct PN field.)
 XX
 SQ Sequence 145 AA;

Query Match 50.9%; Score 299; DB 2; Length 145;
 Best Local Similarity 100.0%; Pred. No. 2.2e-18;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 93 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 145

RESULT 12

AAR71695

ID AAR71695 standard; protein; 146 AA.

XX

AC AAR71695;

XX

DT 25-MAR-2003 (revised)

DT 20-NOV-1995 (first entry)

XX

DE Mating factor alpha 1-Insulin precursor ArgB1, ArgB31 N-terminal.

XX

KW Human insulin precursor ArgB1, ArgB31; diabetes; Zinc ion complex;
 KW mating factor alpha 1; N-terminal EEAEAEAER.
 XX
 OS Homo sapiens.
 XX
 FH Key Location/Qualifiers
 FT Protein 1. .85
 FT /label= mating factor alpha-1
 FT Peptide 86. .94
 FT /label= N-terminal peptide
 FT Peptide 95. .125
 FT /label= B-chain
 FT Peptide 126. .146
 FT /label= A-chain
 XX
 PN WO9507931-A1.
 XX
 PD 23-MAR-1995.
 XX
 PF 16-SEP-1994; 94WO-DK000347.
 XX
 PR 17-SEP-1993; 93DK-00001044.
 PR 02-FEB-1994; 94US-00190829.
 XX
 PA (NOVO) NOVO-NORDISK AS.
 XX
 PI Havelund S, Halstrom JB, Jonassen I, Andersen AS, Markussen J;
 XX
 DR WPI; 1995-131314/17.
 DR N-PSDB; AAQ86432.
 XX
 PT Acylated insulin deriv. which may be present as a Zinc ion complex - is
 PT used to treat diabetes and is rapid acting.
 XX
 PS Example 6; Page 85; 100pp; English.
 XX
 CC AAQ86432 encodes AAR71695 mating factor alpha 1-Insulin precursor ArgB1,
 CC ArgB31 N-terminal EEAEAEAER. The insulin precursor comprises the B and A
 CC chains of a claimed human insulin derivative preceded by the N-terminal
 CC amino acids EEAEAEAER. In the final claimed compsn. they are covalently
 CC connected via disulphide bonds between Cys residues A7/B7 and A20/B19.
 CC The derivative, which may be present as a zinc ion complex, can be used
 CC as a fast action treatment for diabetes. (Updated on 25-MAR-2003 to
 CC correct PN field.)
 XX
 SQ Sequence 146 AA;

Query Match 50.9%; Score 299; DB 2; Length 146;
 Best Local Similarity 100.0%; Pred. No. 2.3e-18;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||||||||||||||||||||||||||
 Db 94 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 146

AA42859

ID AA42859 standard; protein; 52 AA.

XX

AC AA42859;

XX

DT 19-JAN-2000 (first entry)

XX

DE Human insulin precursor, SEQ ID 5.

XX

KW Insulin; precursor; growth hormone; chaperone; intramolecular; folding;
KW conformation; chimeric protein; cleavable; recombinant; production;
KW yield.

XX

OS Homo sapiens.

XX

PN WO9950302-A1.

XX

PD 07-OCT-1999.

XX

PF 31-MAR-1998; 98WO-CN000052.

XX

PR 31-MAR-1998; 98WO-CN000052.

XX

PA (TONG-) TONGHUA GANTECH BIOTECHNOLOGY LTD.

XX

PI Gan Z;

XX

DR WPI; 1999-610839/52.

XX

PT New chimeric proteins containing human growth hormone fragment, used
PT particularly for the production of human insulin.

XX

PS Claim 12; Page 29-30; 46pp; English.

XX

CC This sequence represents a human insulin precursor comprising insulin A
CC and B chains. This insulin precursor is a component of the chimeric
CC proteins hGH-mini-proinsulin (AA42860) and the chimeric protein given in
CC AA42861. These chimeric proteins additionally contain an N-terminal
CC fragment of human growth hormone (hGH) and a cleavable peptide linker
CC (AA42857). The hGH portion of the chimeric protein acts as an
CC intramolecular chaperone (IMC) for the insulin precursor, enabling it to
CC fold correctly. The cleavable peptide linker has a C-terminal Arg residue
CC which enables the hGH portion of the chimeric protein to be removed after
CC folding has taken place. Production of recombinant human insulin via an
CC hGH-proinsulin chimeric protein can provide human insulin with correctly
CC linked cysteine bridges with fewer necessary procedural steps, and hence
CC resulting in a higher yield of human insulin. The IMC sequences not only
CC protect insulin sequences from intracellular degradation by a
CC microorganism host, but also promote the folding of the fused insulin
CC precursor, facilitate the solubility of the fusion protein and decrease
CC the intermolecular interactions among the fusion proteins, thus allowing
CC folding of the fused insulin precursor at commercially useful high
CC concentrations. The procedural steps of cyanogen bromide cleavage,
CC oxidative sulphytolysis and related purification steps can thus be
CC eliminated, along with the use of high concentrations of mercaptan or the
CC use of hydrophobic absorbent resins

XX

SQ Sequence 52 AA;

Query Match 50.1%; Score 294; DB 2; Length 52;
Best Local Similarity 100.0%; Pred. No. 2.4e-18;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|||||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 52

RESULT 14

AAR04582

ID AAR04582 standard; protein; 57 AA.

XX

AC AAR04582;

XX

DT 25-MAR-2003 (revised)

DT 14-SEP-1990 (first entry)

XX

DE Proinsulin analogue with a Lys residue linking the A and B chains.

XX

KW insulin fusion protein; pro-insulin analogue; tendamistate;

KW Lys-Lys bridge; ds.

XX

OS Synthetic.

XX

FH Key Location/Qualifiers

FT Peptide 1. .35

FT /label= Insulin B chain

FT Misc-difference 36

FT /label= Lys residue linking insulin B chain to A chain

FT Peptide 37. .57

FT /label= Insulin A chain

XX

PN EP367163-A.

XX

PD 09-MAY-1990.

XX

PF 28-OCT-1989; 89EP-00120056.

XX

PR 03-NOV-1988; 88DE-03837273.

PR 19-AUG-1989; 89DE-03927449.

XX

PA (FARH) HOECHST AG.

XX

PI Koller KP, Riess GJ, Uhlmann E, Wallmeier H;

XX

DR WPI; 1990-141149/19.

DR N-PSDB; AAQ04335.

XX

PT New insulin fusion proteins - comprise pro-insulin analogue linked to
PT tendamistate.

XX

PS Disclosure; Page ?; -pp; German.

XX

CC This sequence is joined to the C-terminus of an N-terminal fragment

XX
 PS Disclosure; Fig 3; 51pp; English.
 XX
 CC Human recombinant insulin may be expressed in *Saccharomyces cerevisiae*
 CC following linkage of the gene to the glycosylphosphatidylinositol anchor.
 CC This anchoring technique can provide for the release of the product in a
 CC highly specific and selective manner. In addition, the recombinant
 CC protein will contain an epitope which can be used in its final
 CC purification by immunoaffinity. The protein product can be released by
 CC e.g. nitrous deamination or treatment with neutral detergent. (Updated on
 CC 25-MAR-2003 to correct PI field.)
 XX
 SQ Sequence 160 AA;

Query Match 49.1%; Score 288.5; DB 2; Length 160;
 Best Local Similarity 98.1%; Pred. No. 2e-17;
 Matches 53; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKT-RGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||||||||||||| |||||||||||||||||||||
 Db 43 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTKRGIVEQCCTSICSLYQLENYCN 96

Search completed: July 15, 2004, 16:35:35
 Job time : 56.8955 secs

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:30:45 ; Search time 16.1698 Seconds
(without alignments)
341.624 Million cell updates/sec

Title: US-09-423-100-6
Perfect score: 587
Sequence: 1 MFPTIPLSRLFDNAMLRAHR.....IVEQCCTSICSLYQLENYCN 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 389414 seqs, 51625971 residues

Total number of hits satisfying chosen parameters: 389414

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/2/iaa/5A_COMB.pep:*
2: /cgn2_6/ptodata/2/iaa/5B_COMB.pep:*
3: /cgn2_6/ptodata/2/iaa/6A_COMB.pep:*
4: /cgn2_6/ptodata/2/iaa/6B_COMB.pep:*
5: /cgn2_6/ptodata/2/iaa/PCTUS_COMB.pep:*
6: /cgn2_6/ptodata/2/iaa/backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	304	51.8	63	1	US-08-160-376A-6	Sequence 6, Appli
2	302.5	51.5	137	1	US-08-400-256-39	Sequence 39, Appl
3	302.5	51.5	137	3	US-08-975-365-39	Sequence 39, Appl
4	299.5	51.0	66	1	US-08-291-060B-5	Sequence 5, Appli
5	299	50.9	56	1	US-08-160-376A-7	Sequence 7, Appli
6	299	50.9	56	1	US-08-389-487-11	Sequence 11, Appl
7	299	50.9	96	1	US-08-160-376A-5	Sequence 5, Appli
8	299	50.9	96	1	US-08-389-487-8	Sequence 8, Appli
9	299	50.9	145	1	US-08-400-256-45	Sequence 45, Appl
10	299	50.9	145	3	US-08-975-365-45	Sequence 45, Appl
11	299	50.9	146	1	US-08-400-256-48	Sequence 48, Appl

12	299	50.9	146	3	US-08-975-365-48	Sequence 48, Appl
13	293	49.9	57	1	US-08-030-731A-44	Sequence 44, Appl
14	287	48.9	65	3	US-08-900-574-3	Sequence 3, Appli
15	286.5	48.8	66	3	US-08-900-574-5	Sequence 5, Appli
16	286	48.7	67	3	US-08-900-574-7	Sequence 7, Appli
17	284.5	48.5	65	1	US-08-468-674B-71	Sequence 71, Appl
18	284.5	48.5	65	1	US-08-780-571-71	Sequence 71, Appl
19	284.5	48.5	124	3	US-09-012-669F-36	Sequence 36, Appl
20	284	48.4	138	3	US-08-932-082-19	Sequence 19, Appl
21	284	48.4	138	4	US-09-861-687-19	Sequence 19, Appl
22	284	48.4	140	1	US-08-400-256-33	Sequence 33, Appl
23	284	48.4	140	1	US-08-400-256-42	Sequence 42, Appl
24	284	48.4	140	3	US-08-975-365-33	Sequence 33, Appl
25	284	48.4	140	3	US-08-975-365-42	Sequence 42, Appl
26	283.5	48.3	53	1	US-08-233-617-4	Sequence 4, Appli
27	283.5	48.3	53	4	US-08-981-988A-42	Sequence 42, Appl
28	283.5	48.3	117	3	US-09-012-669F-37	Sequence 37, Appl
29	281	47.9	104	1	US-08-400-256-15	Sequence 15, Appl
30	281	47.9	104	3	US-08-975-365-15	Sequence 15, Appl
31	280.5	47.8	89	1	US-08-468-674B-41	Sequence 41, Appl
32	280.5	47.8	89	1	US-08-780-571-41	Sequence 41, Appl
33	280.5	47.8	91	1	US-08-468-674B-45	Sequence 45, Appl
34	280.5	47.8	91	1	US-08-780-571-45	Sequence 45, Appl
35	280.5	47.8	124	1	US-08-446-646-3	Sequence 3, Appli
36	279.5	47.6	167	1	US-07-918-953-8	Sequence 8, Appli
37	279.5	47.6	167	1	US-08-081-661-8	Sequence 8, Appli
38	278.5	47.4	51	4	US-09-477-924-3	Sequence 3, Appli
39	278.5	47.4	51	4	US-09-723-981-3	Sequence 3, Appli
40	278.5	47.4	51	4	US-09-723-896-3	Sequence 3, Appli
41	278	47.4	117	4	US-09-280-030-63	Sequence 63, Appl
42	277.5	47.3	53	1	US-08-233-617-3	Sequence 3, Appli
43	277	47.2	96	2	US-09-134-836-4	Sequence 4, Appli
44	277	47.2	96	4	US-09-386-303A-4	Sequence 4, Appli
45	277	47.2	97	1	US-08-160-376A-4	Sequence 4, Appli

ALIGNMENTS

RESULT 1

US-08-160-376A-6

; Sequence 6, Application US/08160376A

; Patent No. 5473049

; GENERAL INFORMATION:

; APPLICANT: Obermeier, Ranier

; APPLICANT: Gerl, Martin

; APPLICANT: Ludwig, Jurgen

; APPLICANT: Sabel, Walter

; TITLE OF INVENTION: Process For Obtaining Proinsulin

; TITLE OF INVENTION: Possessing Correctly Linked

; TITLE OF INVENTION: Cystine Bridges

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kenneth A. Genoni, Esq.

; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500

; CITY: Somerville

; STATE: New Jersey


```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,256
; FILING DATE: 03-MAR-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-400-256-39

```

```

Query Match          51.5%; Score 302.5; DB 1; Length 137;
Best Local Similarity 50.0%; Pred. No. 3.3e-28;
Matches 70; Conservative 4; Mismatches 27; Indels 39; Gaps 4;

```

```

Qy      2 FPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQ--KYSFLQ-----N 48
      ||:|      | : |      :| |||||      || |:      |
Db      3 FPSI-----FTAVLFAASSALAAPVNTTTTDETAQIPAEAVIGYSDLEGDFDVAVLPEFSN 57

Qy      49 PLGTG-----PRFVNQHLCGSHLVEALYLVCGERGFFYTPKTRG 87
      |      |||||
Db      58 STNGLLFINTTTIASIAAKEEGVSMKRFRVNQHLCGSHLVEALYLVCGERGFFYTPKTRG 117

Qy      88 IVEQCCTSICSLYQLENYCN 107
      |||||
Db      118 IVEQCCTSICSLYQLENYCN 137

```

RESULT 3

US-08-975-365-39

```

; Sequence 39, Application US/08975365
; Patent No. 6011007
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asser Sloth
; APPLICANT: Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York

```



```

; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,365
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/400,256
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-975-365-39

```

```

Query Match          51.5%; Score 302.5; DB 3; Length 137;
Best Local Similarity 50.0%; Pred. No. 3.3e-28;
Matches 70; Conservative 4; Mismatches 27; Indels 39; Gaps 4;

```

```

Qy      2  FPTIPLSRLFDNAMLRAHRLHQLAFDQYEFEEAYIPKEQ--KYSFLQ-----N 48
      ||:|      |      :      |      :|      ||      ||      ||      ||      |
Db      3  FPSI-----FTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVLPFSN 57

Qy      49  PLGTG-----PRFVNQHLCGSHLVEALYLVCGERGFFFYTPKTRG 87
      |      |||
Db      58  STNNGLLFINTTIIASIAAKEEGVSMARFVNQHLCGSHLVEALYLVCGERGFFFYTPKTRG 117

Qy      88  IVEQCCTSICSLYQLENYCN 107
      |||
Db      118 IVEQCCTSICSLYQLENYCN 137

```

RESULT 4

US-08-291-060B-5

; Sequence 5, Application US/08291060B

; Patent No. 5728543

; GENERAL INFORMATION:

; APPLICANT: Dorschug, Michael

; APPLICANT: Koller, Klaus-Peter

; APPLICANT: Marquardt, Rudiger

; APPLICANT: Meiwes, Johannes

; TITLE OF INVENTION: An Enzymatic Process for the

; TITLE OF INVENTION: Conversion of Preproinsulins Into Insulins


```

; ADDRESSEE: Kenneth A. Genoni, Esq.
; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500
; CITY: Somerville
; STATE: New Jersey
; COUNTRY: U.S.A.
; ZIP: 08876-1258
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 386
; OPERATING SYSTEM: WINDOWS 3.1
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/160,376A
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GE P 4240420.7
; FILING DATE: December 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara V. Maurer, Esq.
; REGISTRATION NUMBER: 31,287
; REFERENCE/DOCKET NUMBER: HOE 92/F 384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 231-4079
; TELEFAX: (908) 231-2255
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 56 Amino Acids
; TYPE: Amino Acid (AA)
; TOPOLOGY: not relevant
US-08-160-376A-7

```

```

Query Match          50.9%; Score 299; DB 1; Length 56;
Best Local Similarity 100.0%; Pred. No. 2.8e-28;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|||||
Db      4 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 56

```

RESULT 6

US-08-389-487-11

```

; Sequence 11, Application US/08389487
; Patent No. 5663291
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process for Obtaining Insulin Having
; TITLE OF INVENTION: Correctly Linked Cystine Bridges
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.

```

; CITY: Washington
 ; STATE: D.C.
 ; COUNTRY: United States of America
 ; ZIP: 20005-3315
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/389,487
 ; FILING DATE:
 ; CLASSIFICATION: 530
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Einaudi, Carol P.
 ; REGISTRATION NUMBER: 32,220
 ; REFERENCE/DOCKET NUMBER: 02481.1424-00000
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 202-408-4000
 ; TELEFAX: 202-408-4400
 ; INFORMATION FOR SEQ ID NO: 11:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 56 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: peptide
 US-08-389-487-11

Query Match 50.9%; Score 299; DB 1; Length 56;
 Best Local Similarity 100.0%; Pred. No. 2.8e-28;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 4 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 56

RESULT 7

US-08-160-376A-5

; Sequence 5, Application US/08160376A

; Patent No. 5473049

; GENERAL INFORMATION:

; APPLICANT: Obermeier, Ranier

; APPLICANT: Gerl, Martin

; APPLICANT: Ludwig, Jurgen

; APPLICANT: Sabel, Walter

; TITLE OF INVENTION: Process For Obtaining Proinsulin

; TITLE OF INVENTION: Possessing Correctly Linked

; TITLE OF INVENTION: Cystine Bridges

; NUMBER OF SEQUENCES: 7

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Kenneth A. Genoni, Esq.

; STREET: Rt. 202-206 No. 5473049th/P.O. Box 2500

; CITY: Somerville

; STATE: New Jersey

; COUNTRY: U.S.A.

```

; ZIP: 08876-1258
; COMPUTER READABLE FORM:
; MEDIUM TYPE: DISKETTE, 3.5 INCH, 1.44 Mb STORAGE
; COMPUTER: IBM 386
; OPERATING SYSTEM: WINDOWS 3.1
; SOFTWARE: WORDPERFECT 5.1
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/160,376A
; FILING DATE: December 1, 1993
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: GE P 4240420.7
; FILING DATE: December 2, 1992
; ATTORNEY/AGENT INFORMATION:
; NAME: Barbara V. Maurer, Esq.
; REGISTRATION NUMBER: 31,287
; REFERENCE/DOCKET NUMBER: HOE 92/F 384
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (908) 231-4079
; TELEFAX: (908) 231-2255
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 Amino Acids
; TYPE: Amino Acid (AA)
; TOPOLOGY: not relevant
US-08-160-376A-5

```

```

Query Match          50.9%; Score 299; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 5.5e-28;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
        ||||||||||||||||||||||||||||||||||||||||||||||||
Db      44 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 96

```

```

RESULT 8
US-08-389-487-8
; Sequence 8, Application US/08389487
; Patent No. 5663291
; GENERAL INFORMATION:
; APPLICANT: Obermeier, Rainer
; APPLICANT: Gerl, Martin
; APPLICANT: Ludwig, Jurgen
; APPLICANT: Sabel, Walter
; TITLE OF INVENTION: Process for Obtaining Insulin Having
; TITLE OF INVENTION: Correctly Linked Cystine Bridges
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: D.C.
; COUNTRY: United States of America
; ZIP: 20005-3315
; COMPUTER READABLE FORM:

```

```

; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/389,487
; FILING DATE:
; CLASSIFICATION: 530
; ATTORNEY/AGENT INFORMATION:
; NAME: Einaudi, Carol P.
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1424-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 8:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 96 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: peptide
US-08-389-487-8

```

```

Query Match          50.9%; Score 299; DB 1; Length 96;
Best Local Similarity 100.0%; Pred. No. 5.5e-28;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
        ||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      44 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 96

```

RESULT 9

US-08-400-256-45

; Sequence 45, Application US/08400256

; Patent No. 5750497

; GENERAL INFORMATION:

; APPLICANT: Havelund, Svend

; APPLICANT: Halstrom, John

; APPLICANT: Jonassen, Ib

; APPLICANT: Andersen, Asser Sloth

; APPLICANT: Markussen, Jan

; TITLE OF INVENTION: ACYLATED INSULIN

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

```
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/400,256
; FILING DATE: 03-MAR-1995
; CLASSIFICATION: 514
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 45:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 145 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-400-256-45
```

```
Query Match          50.9%; Score 299; DB 1; Length 145;
Best Local Similarity 100.0%; Pred. No. 9.1e-28;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
Qy      55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
          |||
Db      93 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 145
```

RESULT 10

US-08-975-365-45

```
; Sequence 45, Application US/08975365
; Patent No. 6011007
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; APPLICANT: Halstrom, John
; APPLICANT: Jonassen, Ib
; APPLICANT: Andersen, Asser Sloth
; APPLICANT: Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: No. 6011007o No. 6011007disk of No. 6011007th America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/975,365
; FILING DATE:
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
```

; APPLICATION NUMBER: US 08/400,256
 ; FILING DATE: 03-MAR-1995
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lambiris, Elias J.
 ; REGISTRATION NUMBER: 33,728
 ; REFERENCE/DOCKET NUMBER: 3985.220-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-878-9655
 ; INFORMATION FOR SEQ ID NO: 45:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 145 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 US-08-975-365-45

Query Match 50.9%; Score 299; DB 3; Length 145;
 Best Local Similarity 100.0%; Pred. No. 9.1e-28;
 Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||||||||||||||||||||||||||||||||||
 Db 93 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 145

RESULT 11

US-08-400-256-48

; Sequence 48, Application US/08400256
 ; Patent No. 5750497
 ; GENERAL INFORMATION:
 ; APPLICANT: Havelund, Svend
 ; APPLICANT: Halstrom, John
 ; APPLICANT: Jonassen, Ib
 ; APPLICANT: Andersen, Asser Sloth
 ; APPLICANT: Markussen, Jan
 ; TITLE OF INVENTION: ACYLATED INSULIN
 ; NUMBER OF SEQUENCES: 49
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: No. 5750497o No. 5750497disk of No. 5750497th America, Inc.
 ; STREET: 405 Lexington Avenue, 64th Floor
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/08/400,256
 ; FILING DATE: 03-MAR-1995
 ; CLASSIFICATION: 514
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lambiris, Elias J.
 ; REGISTRATION NUMBER: 33,728

; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 48:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 146 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-975-365-48

Query Match 50.9%; Score 299; DB 3; Length 146;
Best Local Similarity 100.0%; Pred. No. 9.2e-28;
Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|||||
Db 94 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 146

RESULT 13

US-08-030-731A-44

; Sequence 44, Application US/08030731A
; Patent No. 5426036
; GENERAL INFORMATION:
; APPLICANT: Koller, Klaus-Peter
; APPLICANT: Riess, Guenther Johannes
; APPLICANT: Uhlmann, Eugen
; APPLICANT: Wallmeier, Holger
; TITLE OF INVENTION: Processes for the Preparation of Foreign
; TITLE OF INVENTION: Proteins in Streptomyces
; NUMBER OF SEQUENCES: 48
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett &
; ADDRESSEE: Dunner
; STREET: 1300 I Street, N.W., Suite 700
; CITY: Washington
; STATE: D.C.
; COUNTRY: USA
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/030,731A
; FILING DATE: 12-MAR-1993
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/189,840
; FILING DATE: 03-MAY-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/430,622
; FILING DATE: 01-NOV-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/687,610
; FILING DATE: 19-APR-1991

```

; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 07/735,757
; FILING DATE: 29-JUL-1991
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 37 14 866.4
; FILING DATE: 05-MAY-1987
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 38 37 273.8
; FILING DATE: 03-NOV-1988
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 39 27 449.7
; FILING DATE: 19-AUG-1989
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: DE P 40 12 818.0
; FILING DATE: 21-APR-1990
; ATTORNEY/AGENT INFORMATION:
; NAME: Kirschner Michael K.
; REGISTRATION NUMBER: 34,851
; REFERENCE/DOCKET NUMBER: 02481-0593-02000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 202-408-4000
; TELEFAX: 202-408-4400
; INFORMATION FOR SEQ ID NO: 44:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 57 amino acids
; TYPE: amino acid
; TOPOLOGY: unknown
; MOLECULE TYPE: peptide
US-08-030-731A-44

```

```

Query Match          49.9%; Score 293; DB 1; Length 57;
Best Local Similarity 96.2%; Pred. No. 1.5e-27;
Matches 51; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
      :|||||
Db      5  KFVNQHLCGSHLVEALYLVCGERGFFYTPKTKGIVEQCCTSICSLYQLENYCN 57

```

RESULT 14

US-08-900-574-3

; Sequence 3, Application US/08900574

; Patent No. 6221837

; GENERAL INFORMATION:

```

; APPLICANT: Ertl, Johann
; APPLICANT: Habermann, Paul
; APPLICANT: Geisen, Karl
; APPLICANT: Seipke, Gerhard
; TITLE OF INVENTION: Insulin derivatives with increased zinc
; TITLE OF INVENTION: binding
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,
; ADDRESSEE: & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: District of Columbia

```

```

; COUNTRY: U.S.A.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,574
; FILING DATE: July 24, 1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: German Application No. 6221837 19630242.0
; FILING DATE: July 26, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Carol P. Einaudi
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1499-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 65 amino acids
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..65
US-08-900-574-3

```

```

Query Match          48.9%; Score 287; DB 3; Length 65;
Best Local Similarity 91.4%; Pred. No. 8.8e-27;
Matches 53; Conservative 0; Mismatches 3; Indels 2; Gaps 1;

```

```

Qy      51 GTGPRFVNQHLCGSHLVEALYLVCGERGFFYTPKT--RGIVEQCCTSICSLYQLENYC 106
      |  |||
Db      7  GNSARFVNQHLCGSHLVEALYLVCGERGFFYTPKTHHRGIVEQCCTSICSLYQLENYC 64

```

RESULT 15

```

US-08-900-574-5
; Sequence 5, Application US/08900574
; Patent No. 6221837
; GENERAL INFORMATION:
; APPLICANT: Ertl, Johann
; APPLICANT: Habermann, Paul
; APPLICANT: Geisen, Karl
; APPLICANT: Seipke, Gerhard
; TITLE OF INVENTION: Insulin derivatives with increased zinc
; TITLE OF INVENTION: binding
; NUMBER OF SEQUENCES: 18
; CORRESPONDENCE ADDRESS:

```

```

; ADDRESSEE: Finnegan, Henderson, Farabow, Garrett,
; ADDRESSEE: & Dunner, L.L.P.
; STREET: 1300 I Street, N.W.
; CITY: Washington
; STATE: District of Columbia
; COUNTRY: U.S.A.
; ZIP: 20005-3315
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25 (EPO)
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/900,574
; FILING DATE: July 24, 1997
; CLASSIFICATION: 514
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: German Application No. 6221837 19630242.0
; FILING DATE: July 26, 1996
; ATTORNEY/AGENT INFORMATION:
; NAME: Carol P. Einaudi
; REGISTRATION NUMBER: 32,220
; REFERENCE/DOCKET NUMBER: 02481.1499-00000
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (202) 408-4000
; TELEFAX: (202) 408-4400
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 66 amino acids
; TYPE: Amino acid
; STRANDEDNESS: Single
; TOPOLOGY: linear
; MOLECULE TYPE: Protein
; ORIGINAL SOURCE:
; ORGANISM: Escherichia coli
; FEATURE:
; NAME/KEY: Protein
; LOCATION: 1..66
US-08-900-574-5

```

```

Query Match          48.8%; Score 286.5; DB 3; Length 66;
Best Local Similarity 89.8%; Pred. No. 1e-26;
Matches 53; Conservative 0; Mismatches 3; Indels 3; Gaps 1;

```

```

Qy      51 GTGPRFVNQHLCGSHLVEALYLVCGERGFFYTPKT---RGIVEQCCTSICSLYQLENYC 106
|      | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | | |
Db      7  GNSARFVNQHLCGSHLVEALYLVCGERGFFYTPKTAHHRGIVEQCCTSICSLYQLENYC 65

```

```

Search completed: July 15, 2004, 16:42:32
Job time : 16.1698 secs

```

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:29:19 ; Search time 11.9776 Seconds
 (without alignments)
 859.311 Million cell updates/sec

Title: US-09-423-100-6
 Perfect score: 587
 Sequence: 1 MFPTIPLSRLFDNAMLRAHR.....IVEQCCTSICSLYQLENYCN 107

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 283366 seqs, 96191526 residues

Total number of hits satisfying chosen parameters: 283366

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : PIR_78:*
 1: pir1:*
 2: pir2:*
 3: pir3:*
 4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	%		DB	ID	Description
		Query Match	Length			
1	275	46.8	96	2	PC7082	epidermal growth f
2	273.5	46.6	51	1	INWHP	insulin - sperm wh
3	273.5	46.6	51	1	INWHF	insulin - finback
4	273.5	46.6	51	1	INEL	insulin - elephant
5	273	46.5	110	2	B42179	insulin precursor
6	273	46.5	110	2	JQ0178	insulin precursor
7	271.5	46.3	51	1	INHY	insulin - hamster
8	268.5	45.7	51	1	INMSSP	insulin - Egyptian
9	267.5	45.6	51	2	A59151	insulin precursor
10	267	45.5	110	1	IPHU	insulin precursor
11	267	45.5	110	2	A42179	insulin precursor
12	263.5	44.9	51	1	INWH1S	insulin - sei whal
13	263.5	44.9	51	1	INGT	insulin - goat

14	263.5	44.9	51	1	INCMA	insulin - Arabian
15	263	44.8	84	1	IPPG	insulin precursor
16	263	44.8	110	1	INRB	insulin precursor
17	262.5	44.7	51	1	INCT	insulin - cat
18	262	44.6	110	1	IPDG	insulin precursor
19	261.5	44.5	51	1	INMKSQ	insulin - common s
20	260	44.3	110	2	I48166	insulin precursor
21	258.5	44.0	105	1	IPBO	insulin precursor
22	257	43.8	108	2	A39883	insulin precursor
23	256.5	43.7	51	2	JQ0362	insulin - North Am
24	255.5	43.5	217	1	STHU	somatotropin 1 pre
25	255.5	43.5	217	2	I67410	somatotropin - rhe
26	252.5	43.0	77	1	INSH	insulin precursor
27	252	42.9	86	1	IPHO	insulin precursor
28	251.5	42.8	51	1	INCB	insulin - Chinchil
29	250	42.6	108	1	INMS1	insulin 1 precurso
30	249	42.4	110	1	IPRT1	insulin 1 precurso
31	248.5	42.3	51	1	INGS	insulin - goose
32	248	42.2	110	1	IPRT2	insulin 2 precurso
33	248	42.2	110	1	INMS2	insulin 2 precurso
34	246	41.9	52	2	S44469	insulin I1 - North
35	246	41.9	52	2	S44470	insulin I2 - North
36	245	41.7	103	2	I51221	insulin precursor
37	244.5	41.7	51	1	INPQ	insulin - crested
38	244.5	41.7	51	1	INTK	insulin - turkey (
39	244.5	41.7	51	1	INOS	insulin - ostrich
40	244.5	41.7	51	1	A61129	insulin - black-be
41	244.5	41.7	51	2	A60414	insulin - slider t
42	239.5	40.8	107	1	IPCH	insulin precursor
43	238	40.5	52	2	S61361	insulin - Amphiuma
44	235.5	40.1	51	2	S63590	insulin - duckbill
45	233.5	39.8	81	1	IPDK	insulin precursor

ALIGNMENTS

RESULT 1

PC7082

epidermal growth factor/single chain insulin fusion protein - Bacillus brevis
(fragment)

C;Species: Bacillus brevis

C;Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 31-Mar-2003

C;Accession: PC7082; PC7083

R;Koh, M.; Hanagata, H.; Ebisu, S.; Morihara, K.; Takagi, H.

Biosci. Biotechnol. Biochem. 64, 1079-1081, 2000

A;Title: Use of Bacillus brevis for synthesis and secretion of Des-B30 single-chain human insulin precursor.

A;Reference number: PC7082; MUID:20335834; PMID:10879487

A;Accession: PC7082

A;Molecule type: DNA

A;Residues: 1-96 <KOH>

A;Accession: PC7083

A;Molecule type: protein

A;Residues: 19-28 <KO2>

C;Genetics:

A;Gene: egf-sci

C;Superfamily: insulin

Query Match 46.8%; Score 275; DB 2; Length 96;
Best Local Similarity 94.3%; Pred. No. 1.2e-21;
Matches 50; Conservative 1; Mismatches 0; Indels 2; Gaps 1;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
:|||||
Db 46 KFVNQHLCGSHLVEALYLVCGERGFFYTPK--GIVEQCCTSICSLYQLENYCN 96

RESULT 2

INWHP

insulin - sperm whale

C;Species: Physeter catodon (sperm whale)

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999

C;Accession: A93142; A90082

R;Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.

Nature 181, 1468-1469, 1958

A;Title: Structure of sperm- and sei-whale insulins and their breakdown by whale pepsin.

A;Reference number: A93142

A;Accession: A93142

A;Molecule type: protein

A;Residues: 1-30;31-51 <ISH>

R;Harris, J.I.; Sanger, F.; Naughton, M.A.

Arch. Biochem. Biophys. 65, 427-428, 1956

A;Title: Species differences in insulin.

A;Reference number: A90082

A;Accession: A90082

A;Molecule type: protein

A;Residues: 1-30;31-51 <HAR>

C;Superfamily: insulin

C;Keywords: hormone; pancreas

F;1-30/Domain: insulin chain B #status experimental <BCH>

F;1-30,31-51/Product: insulin #status experimental <MAT>

F;31-51/Domain: insulin chain A #status experimental <ACH>

F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 46.6%; Score 273.5; DB 1; Length 51;
Best Local Similarity 96.2%; Pred. No. 8.7e-22;
Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|||||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSICSLYQLENYCN 51

RESULT 3

INWHF

insulin - finback whale (tentative sequence)

C;Species: Balaenoptera physalus (finback whale, common rorqual)

C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999

C;Accession: A91918

R;Hama, H.; Titani, K.; Sakaki, S.; Narita, K.

J. Biochem. 56, 285-293, 1964

A;Title: The amino acid sequence in fin-whale insulin.

A;Reference number: A91918
A;Accession: A91918
A;Molecule type: protein
A;Residues: 1-30;31-51 <HAM>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCH>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
| | | | | | | | | | | | | | | | | | | | | | |
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKA-GIVEOCCTSICSLYOLENYCN 51

Query Match 46.6%; Score 273.5; DB 1; Length 51;
Best Local Similarity 94.2%; Pred. No. 8.7e-22;
Matches 49; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|||||
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTGVCSLYOLENYCN 51

C;Superfamily: insulin
F;1-24/Domain: signal sequence #status predicted <SIG>
F;25-54,90-110/Product: insulin #status predicted <MAT>
F;25-54/Domain: insulin chain B #status predicted <BCH>
F;55-89/Domain: insulin connecting C peptide #status predicted <CPT>
F;90-110/Domain: insulin chain A #status predicted <ACH>
F;31-96,43-109,95-100/Disulfide bonds: #status predicted

Query Match 46.5%; Score 273; DB 2; Length 110;
Best Local Similarity 60.2%; Pred. No. 2.2e-21;
Matches 53; Conservative 0; Mismatches 1; Indels 34; Gaps 1;

```
Qy      54 PRFVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
          | |||||
Db      23 PAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLAL 82

Qy      86 -----RGIVEQCCTSICSLYQLENYCN 107
          |||||
Db      83 EGS LQKRGIVEQCCTSICSLYQLENYCN 110
```

RESULT 7

INHY
insulin - hamster
C;Species: Cricetinae gen. sp. (hamster)
C;Date: 31-Mar-1992 #sequence_revision 31-Mar-1992 #text_change 16-Jul-1999
C;Accession: A91456
R;Neelon, F.A.; Delcher, H.K.; Steinman, H.; Lebovitz, H.E.
Fed. Proc. 32, 300, 1973
A;Title: Structure of hamster insulin: comparison with a tumor insulin.
A;Reference number: A91456
A;Accession: A91456
A;Molecule type: protein
A;Residues: 1-30;31-51 <NEE>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCH>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 46.3%; Score 271.5; DB 1; Length 51;
Best Local Similarity 94.2%; Pred. No. 1.4e-21;
Matches 49; Conservative 2; Mismatches 0; Indels 1; Gaps 1;

```
Qy      56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
          |||||: |||: |||||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSICSLYQLENYCN 51
```

RESULT 8

INMSSP
insulin - Egyptian spiny mouse (tentative sequence)
C;Species: Acomys cahirinus (Egyptian spiny mouse)
C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 31-Mar-2000
C;Accession: A01591
R;Buenzli, H.F.; Humbel, R.E.

Hoppe-Seyler's Z. Physiol. Chem. 353, 444-450, 1972
A;Title: Isolation and partial structural analysis of insulin from mouse (Mus musculus) and spiny mouse (Acomys cahirinus).
A;Reference number: A01591; MUID:72189454; PMID:5028210
A;Contents: composition
A;Accession: A01591
A;Molecule type: protein
A;Residues: 1-30;31-51 <BUE>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status predicted <BCH>
F;1-30,31-51/Product: insulin #status predicted <MAT>
F;31-51/Domain: insulin chain A #status predicted <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 45.7%; Score 268.5; DB 1; Length 51;
Best Local Similarity 92.3%; Pred. No. 2.9e-21;
Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|:|||||||||||||||||||: |||:|||||||||||||||
Db 1 FVBQHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSICSLYQLENYCN 51

RESULT 9

A59151
insulin precursor - jack bean (fragments)
N;Alternate names: hypoglycemic agent; plant insulin
C;Species: Canavalia ensiformis (jack bean)
C;Date: 07-Dec-1999 #sequence_revision 07-Dec-1999 #text_change 10-Dec-1999
C;Accession: B59151; A59151
R;Oliveira, A.E.A.; Machado, O.L.T.; Gomes, V.M.; Xavier-Neto, J.; Pereira, A.C.P.; Vieira, J.G.H.; Fernandes, K.V.S.; Xavier-Filho, J.
Protein Pept. Lett. 6, 15-21, 1999
A;Title: Jack bean seed coat contains a protein with complete sequence homology to bovine insulin.
A;Reference number: A59151
A;Accession: B59151
A;Molecule type: protein
A;Residues: 1-30 <MACB>
A;Accession: A59151
A;Molecule type: protein
A;Residues: 31-51 <MACA>
C;Comment: The two chains are probably produced from the same precursor.
C;Superfamily: insulin
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;1-30/Domain: chain B #status experimental <CHB>
F;31-51/Domain: chain A #status experimental <CHA>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Query Match 45.6%; Score 267.5; DB 2; Length 51;
Best Local Similarity 92.3%; Pred. No. 3.7e-21;
Matches 48; Conservative 1; Mismatches 2; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
||||||||||||||||||| ||||| |:|||||||||||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCASVCSLYQLENYCN 51

RESULT 10

IPHU

insulin precursor [validated] - human

N;Alternate names: preproinsulin

C;Species: Homo sapiens (man)

C;Date: 23-Oct-1981 #sequence_revision 23-Oct-1981 #text_change 08-Dec-2000

C;Accession: A93222; A94253; A93216; A94251; A93144; A92075; A91186; I58114; A01579; S58661

R;Bell, G.I.; Pictet, R.L.; Rutter, W.J.; Cordell, B.; Tischer, E.; Goodman, H.M.

Nature 284, 26-32, 1980

A;Title: Sequence of the human insulin gene.

A;Reference number: A93222; MUID:80120725; PMID:6243748

A;Accession: A93222

A;Molecule type: DNA

A;Residues: 1-110 <BEL>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Ullrich, A.; Dull, T.J.; Gray, A.; Brosius, J.; Sures, I.

Science 209, 612-615, 1980

A;Title: Genetic variation in the human insulin gene.

A;Reference number: A94253; MUID:80236313; PMID:6248962

A;Accession: A94253

A;Molecule type: DNA

A;Residues: 1-110 <ULL>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Bell, G.I.; Swain, W.F.; Pictet, R.; Cordell, B.; Goodman, H.M.; Rutter, W.J.

Nature 282, 525-527, 1979

A;Title: Nucleotide sequence of a cDNA clone encoding human preproinsulin.

A;Reference number: A93216; MUID:80054779; PMID:503234

A;Accession: A93216

A;Molecule type: mRNA

A;Residues: 1-110 <BEL2>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Sures, I.; Goeddel, D.V.; Gray, A.; Ullrich, A.

Science 208, 57-59, 1980

A;Title: Nucleotide sequence of human preproinsulin complementary DNA.

A;Reference number: A94251; MUID:80147417; PMID:6927840

A;Accession: A94251

A;Molecule type: mRNA

A;Residues: 1-110 <SUR>

A;Cross-references: GB:J00265; NID:g186429; PIDN:AAA59172.1; PID:g386828

R;Nicol, D.S.H.W.; Smith, L.F.

Nature 187, 483-485, 1960

A;Title: Amino-acid sequence of human insulin.

A;Reference number: A93144

A;Accession: A93144

A;Molecule type: protein

A;Residues: 25-54;90-110 <NIC>

R;Oyer, P.E.; Cho, S.; Peterson, J.D.; Steiner, D.F.

J. Biol. Chem. 246, 1375-1386, 1971

A;Title: Studies on human proinsulin. Isolation and amino acid sequence of the human pancreatic C-peptide.

A;Reference number: A92075; MUID:71116410; PMID:5101771

A;Accession: A92075

A;Molecule type: protein

A;Residues: 57-87 <OYE>
 R;Ko, A.; Smyth, D.G.; Markussen, J.; Sundby, F.
 Eur. J. Biochem. 20, 190-199, 1971
 A;Title: Amino acid sequence of the C-peptide of human proinsulin.
 A;Reference number: A91186; MUID:71257722; PMID:5560404
 A;Accession: A91186
 A;Molecule type: protein
 A;Residues: 57-87 <KOA>
 R;Lucassen, A.M.; Julier, C.; Beressi, J.P.; Boitard, C.; Froguel, P.; Lathrop, M.; Bell, J.I.
 Nature Genet. 4, 305-310, 1993
 A;Title: Susceptibility to insulin dependent diabetes mellitus maps to a 4.1 kb segment of DNA spanning the insulin gene and associated VNTR.
 A;Reference number: I58114; MUID:93364428; PMID:8358440
 A;Accession: I58114
 A;Status: preliminary; translated from GB/EMBL/DDBJ
 A;Molecule type: DNA
 A;Residues: 1-59,63-110 <RES>
 A;Cross-references: GB:L15440; NID:g307071; PIDN:AAA59179.1; PID:g307072
 R;Sieber, P.; Kamber, B.; Hartmann, A.; Joehl, A.; Riniker, B.; Rittel, W.
 Helv. Chim. Acta 57, 2617-2621, 1974
 A;Title: Totalsynthese von Humaninsulin unter gezielter Bildung der Disulfidbindungen.
 A;Reference number: A91636; MUID:75077277; PMID:4443293
 A;Contents: annotation; synthesis
 A;Note: disulfide-bonded human insulin was synthesized; the synthetic hormone was identical with the natural hormone in chemical and biological activities
 A;Note: article in German with English abstract
 R;Naithani, V.K.
 Hoppe-Seyler's Z. Physiol. Chem. 354, 659-672, 1973
 A;Title: The synthesis of C-peptide of human proinsulin.
 A;Reference number: A91658; MUID:75040007; PMID:4803504
 A;Contents: annotation; synthesis of residues 57-87
 R;Geiger, R.; Jaeger, G.; Koenig, W.
 Chem. Ber. 106, 2347-2352, 1973
 A;Title: Synthesis of the complete sequence of human proinsulin C-peptide and its [Glu-9,Gln-11] analogue.
 A;Reference number: A90914
 A;Contents: annotation; synthesis of residues 57-87
 R;Kaufmann, J.E.; Irminger, J.C.; Halban, P.A.
 Biochem. J. 310, 869-874, 1995
 A;Title: Sequence requirements for proinsulin processing at the B-chain/C-peptide junction.
 A;Reference number: S58661; MUID:96013185; PMID:7575420
 A;Contents: annotation; site-directed mutagenesis study of proteolytic processing
 C;Genetics:
 A;Gene: GDB:INS
 A;Cross-references: GDB:119349; OMIM:176730
 A;Map position: 11p15.5-11p15.5
 A;Introns: 63/1
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-24/Domain: signal sequence #status predicted <SIG>
 F;25-54/Domain: insulin chain B #status experimental <BCH>
 F;25-54,90-110/Product: insulin #status experimental <MAT>
 F;57-87/Domain: connecting C peptide #status experimental <CPEP>

F;90-110/Domain: insulin chain A #status experimental <ACH>
F;31-96,43-109,95-100/Disulfide bonds: #status experimental

Query Match 45.5%; Score 267; DB 1; Length 110;
Best Local Similarity 60.5%; Pred. No. 9.3e-21;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

```
Qy      56 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      86 ----RGIVEQCCTSICSLYQLENYCN 107
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

RESULT 11

A42179

insulin precursor - chimpanzee

C;Species: Pan troglodytes (chimpanzee)

C;Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 16-Jul-1999

C;Accession: A42179; S22058

R;Seino, S.; Bell, G.I.; Li, W.H.

Mol. Biol. Evol. 9, 193-203, 1992

A;Title: Sequences of primate insulin genes support the hypothesis of a slower rate of molecular evolution in humans and apes than in monkeys.

A;Reference number: A42179; MUID:92219953; PMID:1560757

A;Accession: A42179

A;Status: preliminary

A;Molecule type: DNA

A;Residues: 1-110 <SET>

A;Cross-references: EMBL:X61089; NID:g38251; PIDN:CAA43403.1; PID:g38252

A;Note: sequence extracted from NCBI backbone (NCBIP:95067)

C;Genetics:

A;Introns: 63/1

C;Superfamily: insulin

Query Match 45.5%; Score 267; DB 2; Length 110;
Best Local Similarity 60.5%; Pred. No. 9.3e-21;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

```
Qy      56 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
          |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      86 ----RGIVEQCCTSICSLYQLENYCN 107
          |||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110
```

RESULT 12

INWH1S

insulin - sei whale

C;Species: Balaenoptera borealis (sei whale)

C;Date: 13-Jul-1981 #sequence_revision 13-Jul-1981 #text_change 16-Jul-1999

C;Accession: A01582

R;Ishihara, Y.; Saito, T.; Ito, Y.; Fujino, M.

C;Accession: A92782
R;Danho, W.O.
J. Fac. Med. Baghdad 14, 16-28, 1972
A;Title: The isolation and characterization of insulin of camel (*Camelus dromedarius*).
A;Reference number: A92782
A;Accession: A92782
A;Molecule type: protein
A;Residues: 1-30;31-51 <DAN>
C;Superfamily: insulin
C;Keywords: hormone; pancreas
F;1-30/Domain: insulin chain B #status experimental <BCH>
F;1-30,31-51/Product: insulin #status experimental <MAT>
F;31-51/Domain: insulin chain A #status experimental <ACH>
F;7-37,19-50,36-41/Disulfide bonds: #status predicted

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
| | | | | | | | | | | | | | | | | | | | | | : | | | | |
Db 1 FANQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCASVCSLYOLENYCN 51

A;Residues: 33-38,40-62 <SNE>
 A;Note: the authors report the characterization of a connecting peptide variant lacking Ala-39
 A;Accession: B60835
 A;Molecule type: protein
 A;Residues: 33-62 <SN2>
 R;Blundell, T.; Dodson, G.; Hodgkin, D.; Mercola, D.
 Adv. Protein Chem. 26, 279-402, 1972
 A;Title: Insulin. the structure in the crystal and its reflection in chemistry and biology.
 A;Reference number: A90017
 A;Contents: annotation; X-ray crystallography, 1.9 angstroms
 C;Superfamily: insulin
 C;Keywords: hormone; pancreas
 F;1-30/Domain: insulin chain B #status experimental <BCH>
 F;1-30,64-84/Product: insulin #status experimental <MAT>
 F;33-63/Domain: connecting peptide #status experimental <CPEP>
 F;64-84/Domain: insulin chain A #status experimental <ACH>
 F;7-70,19-83,69-74/Disulfide bonds: #status experimental

Query Match 44.8%; Score 263; DB 1; Length 84;
 Best Local Similarity 60.7%; Pred. No. 1.8e-20;
 Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;

```

Qy      56 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
          |||||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAENPQAGAVELGGGLGGLQALALEGPP 60

Qy      86 --RGIVEQCCTSIQSLYQLENYCN 107
          |||||
Db      61 QKRGIVEQCCTSIQSLYQLENYCN 84
  
```

Search completed: July 15, 2004, 16:37:33
 Job time : 12.1443 secs

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:37:41 ; Search time 44.5168 Seconds
(without alignments)
751.267 Million cell updates/sec

Title: US-09-423-100-6
Perfect score: 587
Sequence: 1 MFPTIPLSRLFDNAMLRAHR.....IVEQCCTSICSLYQLENYCN 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1285345 seqs, 312560633 residues

Total number of hits satisfying chosen parameters: 1285345

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09A_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09B_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US09C_PUBCOMB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US10A_PUBCOMB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US10B_PUBCOMB.pep:*
15: /cgn2_6/ptodata/1/pubpaa/US10C_PUBCOMB.pep:*
16: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
17: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
18: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

8

Result	Query				
No.	Score	Match	Length	ID	Description

1	587	100.0	107	13	US-10-054-873-6	Sequence 6, Appli
2	555.5	94.6	150	13	US-10-054-873-7	Sequence 7, Appli
3	302.5	51.5	137	16	US-10-101-454-39	Sequence 39, Appl
4	299	50.9	145	16	US-10-101-454-45	Sequence 45, Appl
5	299	50.9	146	16	US-10-101-454-48	Sequence 48, Appl
6	294	50.1	52	13	US-10-054-873-5	Sequence 5, Appli
7	284.5	48.5	124	9	US-09-894-711-18	Sequence 18, Appl
8	284	48.4	138	9	US-09-861-687-19	Sequence 19, Appl
9	284	48.4	138	12	US-10-620-651-19	Sequence 19, Appl
10	284	48.4	140	16	US-10-101-454-33	Sequence 33, Appl
11	284	48.4	140	16	US-10-101-454-42	Sequence 42, Appl
12	281	47.9	104	16	US-10-101-454-15	Sequence 15, Appl
13	278.5	47.4	51	10	US-09-858-935B-5	Sequence 5, Appli
14	278.5	47.4	51	12	US-10-444-649-3	Sequence 3, Appli
15	278.5	47.4	51	12	US-10-444-701-3	Sequence 3, Appli
16	278.5	47.4	51	12	US-10-271-869-5	Sequence 5, Appli
17	278.5	47.4	51	13	US-10-028-410-3	Sequence 3, Appli
18	278.5	47.4	51	14	US-10-444-326-3	Sequence 3, Appli
19	278.5	47.4	51	16	US-10-444-262-3	Sequence 3, Appli
20	278	47.4	117	9	US-09-280-030-63	Sequence 63, Appl
21	277.5	47.3	124	15	US-10-221-677-24	Sequence 24, Appl
22	277	47.2	96	9	US-09-947-563-4	Sequence 4, Appli
23	277	47.2	102	16	US-10-101-454-36	Sequence 36, Appl
24	275.5	46.9	124	9	US-09-736-611-12	Sequence 12, Appl
25	275.5	46.9	124	9	US-09-740-359-12	Sequence 12, Appl
26	275.5	46.9	124	9	US-09-894-711-12	Sequence 12, Appl
27	275.5	46.9	124	14	US-10-316-421-12	Sequence 12, Appl
28	275.5	46.9	125	9	US-09-736-611-10	Sequence 10, Appl
29	275.5	46.9	125	9	US-09-740-359-10	Sequence 10, Appl
30	275.5	46.9	125	9	US-09-894-711-10	Sequence 10, Appl
31	275.5	46.9	125	14	US-10-316-421-10	Sequence 10, Appl
32	275.5	46.9	147	9	US-09-736-611-8	Sequence 8, Appli
33	275.5	46.9	147	9	US-09-740-359-7	Sequence 7, Appli
34	275.5	46.9	147	14	US-10-316-421-8	Sequence 8, Appli
35	274	46.7	144	9	US-09-736-611-6	Sequence 6, Appli
36	274	46.7	144	9	US-09-740-359-5	Sequence 5, Appli
37	274	46.7	144	14	US-10-316-421-6	Sequence 6, Appli
38	274	46.7	146	9	US-09-894-711-5	Sequence 5, Appli
39	273	46.5	50	13	US-10-066-009A-3	Sequence 3, Appli
40	271	46.2	96	9	US-09-947-563-5	Sequence 5, Appli
41	270	46.0	104	16	US-10-101-454-21	Sequence 21, Appl
42	270	46.0	104	16	US-10-101-454-27	Sequence 27, Appl
43	269.5	45.9	130	9	US-09-280-030-62	Sequence 62, Appl
44	269	45.8	104	16	US-10-101-454-24	Sequence 24, Appl
45	269	45.8	104	16	US-10-101-454-30	Sequence 30, Appl

ALIGNMENTS

RESULT 1

US-10-054-873-6

; Sequence 6, Application US/10054873

; Publication No. US20020164712A1

; GENERAL INFORMATION:

; APPLICANT: Gan, Zhong Ru

```

; TITLE OF INVENTION: Chimeric Protein Containing an
;                      Intramolecular Chaperone-Like Sequence
;
; NUMBER OF SEQUENCES: 7
;
; CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Townsend and Townsend and Crew LLP
;     STREET: Two Embarcadero Center, Eighth Floor
;     CITY: San Francisco
;     STATE: California
;     COUNTRY: USA
;     ZIP: 94111-3834
;
; COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
;
; CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/054,873
;     FILING DATE: 22-Jan-2002
;     CLASSIFICATION: <Unknown>
;
; PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: WO PCT/CN98/00052
;     FILING DATE: 31-MAR-1998
;     APPLICATION NUMBER: US 09/423,100
;     FILING DATE: 11-DEC-2000
;
; ATTORNEY/AGENT INFORMATION:
;     NAME: Mycroft, Frank J
;     REGISTRATION NUMBER: 46,946
;     REFERENCE/DOCKET NUMBER: 020167-000130US
;
; INFORMATION FOR SEQ ID NO: 6:
;     SEQUENCE CHARACTERISTICS:
;         LENGTH: 107 amino acids
;         TYPE: amino acid
;         STRANDEDNESS: <Unknown>
;         TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-10-054-873-6

```

```

Query Match          100.0%; Score 587; DB 13; Length 107;
Best Local Similarity 100.0%; Pred. No. 4.8e-62;
Matches 107; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```

```

Qy      1 MFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQKYSFLQNPLGTGPRFVNQH 60
        ||||||||||||||||||||||||||||||||||||||||||||||||||||||||
Db      1 MFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQKYSFLQNPLGTGPRFVNQH 60

Qy     61 LCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
        ||||||||||||||||||||||||||||||||||||||||||||
Db     61 LCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107

```

```

RESULT 2
US-10-054-873-7
; Sequence 7, Application US/10054873
; Publication No. US20020164712A1
; GENERAL INFORMATION:
;     APPLICANT: Gan, Zhong Ru

```

```

; TITLE OF INVENTION: Chimeric Protein Containing an
;                      Intramolecular Chaperone-Like Sequence
; NUMBER OF SEQUENCES: 7
; CORRESPONDENCE ADDRESS:
;     ADDRESSEE: Townsend and Townsend and Crew LLP
;     STREET: Two Embarcadero Center, Eighth Floor
;     CITY: San Francisco
;     STATE: California
;     COUNTRY: USA
;     ZIP: 94111-3834
; COMPUTER READABLE FORM:
;     MEDIUM TYPE: Floppy disk
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/054,873
;     FILING DATE: 22-Jan-2002
;     CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: WO PCT/CN98/00052
;     FILING DATE: 31-MAR-1998
;     APPLICATION NUMBER: US 09/423,100
;     FILING DATE: 11-DEC-2000
; ATTORNEY/AGENT INFORMATION:
;     NAME: Mycroft, Frank J
;     REGISTRATION NUMBER: 46,946
;     REFERENCE/DOCKET NUMBER: 020167-000130US
; INFORMATION FOR SEQ ID NO: 7:
;     SEQUENCE CHARACTERISTICS:
;     LENGTH: 150 amino acids
;     TYPE: amino acid
;     STRANDEDNESS: <Unknown>
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     SEQUENCE DESCRIPTION: SEQ ID NO: 7:
US-10-054-873-7

```

```

Query Match          94.6%;  Score 555.5;  DB 13;  Length 150;
Best Local Similarity 71.3%;  Pred. No. 4.2e-58;
Matches 107;  Conservative 0;  Mismatches 0;  Indels 43;  Gaps 1;

```

```

Qy      1 MFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQKYSFLQNP----- 49
          ||||||||||||||||||||||||||||||||||||||||||||
Db      1 MFPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQKYSFLQNPQTSLSFSSESIP 60

Qy     50 -----LGTGPRFVNQHLCGSHLVEALYLVCGER 77
          ||||||||||||||||||||||||||||||||||||||||
Db     61 TPSNREETQQKSNLELLRISLLLIQSWLEPVQLGTGPRFVNQHLCGSHLVEALYLVCGER 120

Qy     78 GFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
          ||||||||||||||||||||||||||||
Db    121 GFFYTPKTRGIVEQCCTSICSLYQLENYCN 150

```

RESULT 3
US-10-101-454-39

```

; Sequence 39, Application US/10101454
; Publication No. US20040110664A1
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; Halstrom, John
; Jonassen, Ib
; Andersen, Asser Sloth
; Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/101,454
; FILING DATE: 20-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/400,256
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 39:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 137 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 39:
US-10-101-454-39

```

```

Query Match          51.5%; Score 302.5; DB 16; Length 137;
Best Local Similarity 50.0%; Pred. No. 6e-28;
Matches 70; Conservative 4; Mismatches 27; Indels 39; Gaps 4;

```

```

Qy      2  FPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQ--KYSFLQ-----N 48
      ||:|      |      :      |      |      |      |      |      |
Db      3  FPSI-----FTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLGDFDVAVL PFSN 57

Qy      49  PLGTG-----PRFVNQHLCGSHLVEALYLVCGERGFFYTPKTRG 87
      |      |      |      |      |      |      |      |      |
Db      58  STNNGLLFINTTIIASIAAKEEGVSMARFVNQHLCGSHLVEALYLVCGERGFFYTPKTRG 117

```

Qy 88 IVEQCCTSICSLYQLENYCN 107
|||||
Db 118 IVEQCCTSICSLYQLENYCN 137

RESULT 4

US-10-101-454-45

; Sequence 45, Application US/10101454

; Publication No. US20040110664A1

; GENERAL INFORMATION:

; APPLICANT: Havelund, Svend

; Halstrom, John

; Jonassen, Ib

; Andersen, Asser Sloth

; Markussen, Jan

; TITLE OF INVENTION: ACYLATED INSULIN

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Novo Nordisk of North America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/101,454

; FILING DATE: 20-Mar-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/400,256

; FILING DATE: 03-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 3985.220-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 45:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 145 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 45:

US-10-101-454-45

Query Match 50.9%; Score 299; DB 16; Length 145;

Best Local Similarity 100.0%; Pred. No. 1.7e-27;

Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107

Db 93 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 145

RESULT 5

US-10-101-454-48

; Sequence 48, Application US/10101454

; Publication No. US20040110664A1

; GENERAL INFORMATION:

; APPLICANT: Havelund, Svend

; Halstrom, John

; Jonassen, Ib

; Andersen, Asser Sloth

; Markussen, Jan

; TITLE OF INVENTION: ACYLATED INSULIN

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Novo Nordisk of North America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/101,454

; FILING DATE: 20-Mar-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/400,256

; FILING DATE: 03-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 3985.220-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 48:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 146 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 48:

US-10-101-454-48

Query Match 50.9%; Score 299; DB 16; Length 146;

Best Local Similarity 100.0%; Pred. No. 1.7e-27;

Matches 53; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 55 RFVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107

|||||

US-10-054-873-5

US-10-054-873-5

Query Match 50.1%; Score 294; DB 13; Length 52;
Best Local Similarity 100.0%; Pred. No. 1.9e-27;
Matches 52; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSI~~C~~SLYQLENYCN 107
 |||||
Db 1 FVNOHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSI~~C~~SLYQLENYCN 52

RESULT 7

US-09-894-711-18
 ; Sequence 18, Application US/09894711
 ; Patent No. US20020137144A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Kjeldsen, Thomas Borglum
 ; APPLICANT: Ludvigsen, Svend
 ; TITLE OF INVENTION: Method for making insulin precursors and
 ; TITLE OF INVENTION: insulin precursor analogues having improved fermentation
 ; TITLE OF INVENTION: yield in yeast
 ; FILE REFERENCE: 6148.400-US
 ; CURRENT APPLICATION NUMBER: US/09/894,711
 ; CURRENT FILING DATE: 2001-06-28
 ; PRIOR APPLICATION NUMBER: PA 2000 00443
 ; PRIOR FILING DATE: 2000-03-17
 ; PRIOR APPLICATION NUMBER: PA 1999 01869
 ; PRIOR FILING DATE: 1999-12-29
 ; PRIOR APPLICATION NUMBER: 60/211,081
 ; PRIOR FILING DATE: 2000-06-13
 ; PRIOR APPLICATION NUMBER: 60/181,450
 ; PRIOR FILING DATE: 2000-02-10
 ; PRIOR APPLICATION NUMBER: 09/740,359
 ; PRIOR FILING DATE: 2000-12-19
 ; NUMBER OF SEQ ID NOS: 20
 ; SOFTWARE: FastSEQ for Windows Version 4.0
 ; SEQ ID NO 18
 ; LENGTH: 124
 ; TYPE: PRT
 ; ORGANISM: Artificial Sequence
 ; FEATURE:
 ; OTHER INFORMATION: Synthetic
 US-09-894-711-18

Query Match 48.5%; Score 284.5; DB 9; Length 124;
 Best Local Similarity 92.7%; Pred. No. 7.4e-26;
 Matches 51; Conservative 2; Mismatches 1; Indels 1; Gaps 1;

Qy 54 PRFVNQHLGSHLVEALYLVCGERGFFYTPK-TRGIVEQCCTSICSLYQLENYCN 107
 |:|||||:|||||
 Db 70 PKFVNQHLGSHLVEALYLVCGERGFFYTPKAAKGIVEQCCTSICSLYQLENYCN 124

RESULT 8
 US-09-861-687-19
 ; Sequence 19, Application US/09861687
 ; Publication No. US20020193292A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Markussen, Jan
 ; Jonassen, Ib
 ; Havelund, Svend
 ; Brandt, Jakob
 ; Kurtzhals, Peter
 ; Hansen, Hertz Per
 ; Kaarsholm, Niels Christian
 ; TITLE OF INVENTION: INSULIN DERIVATIVES
 ; NUMBER OF SEQUENCES: 26
 ; CORRESPONDENCE ADDRESS:

; ADDRESSEE: No. US20020193292A1o No. US20020193292A1disk of No.
 US20020193292A1th America, Inc.
 ; STREET: 405 Lexington Avenue, 64th Floor
 ; CITY: New York
 ; STATE: New York
 ; COUNTRY: United States of America
 ; ZIP: 10174-6401
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.30
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/861,687
 ; FILING DATE: 21-May-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/932,082
 ; FILING DATE: 16-DEC-1997
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Lambiris, Elias J.
 ; REGISTRATION NUMBER: 33,728
 ; REFERENCE/DOCKET NUMBER: 4341.204-US
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: 212-867-0123
 ; TELEFAX: 212-878-9655
 ; INFORMATION FOR SEQ ID NO: 19:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 138 amino acids
 ; TYPE: amino acid
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 19:
 US-09-861-687-19

Query Match 48.4%; Score 284; DB 9; Length 138;
 Best Local Similarity 48.2%; Pred. No. 9.8e-26;
 Matches 68; Conservative 5; Mismatches 28; Indels 40; Gaps 5;

Qy 2 FPTIPLSRLEFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQ--KYSFLQ-----N 48
 ||:| | : | :| ||||| | |||:
 Db 3 FPSI-----FTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGFDFVAVLPFSN 57
 Qy 49 PLGTG-----PRFVNQHLCGSHLVEALYLVCGERGFFYTPK-TR 86
 | ||||| :
 Db 58 STNNGLLFINTTIIASIAAKEEGVSLDKRFVNQHLCGSHLVEALYLVCGERGFFYTPKAAK 117
 Qy 87 GIVEQCCTSICSLYQLENYCN 107
 |||||
 Db 118 GIVEQCCTSICSLYQLENYCN 138

RESULT 9
 US-10-620-651-19
 ; Sequence 19, Application US/10620651
 ; Publication No. US20040067874A1
 ; GENERAL INFORMATION:

```

;      APPLICANT: Markussen, Jan
;              Jonassen, Ib
;              Havelund, Svend
;              Brandt, Jakob
;              Kurtzhals, Peter
;              Hansen, Hertz Per
;              Kaarsholm, Niels Christian
;      TITLE OF INVENTION: INSULIN DERIVATIVES
;      NUMBER OF SEQUENCES: 26
;      CORRESPONDENCE ADDRESS:
;              ADDRESSEE: No. US20040067874Alo No. US20040067874Aldisk of No.
US20040067874Aalth America, Inc.
;              STREET: 405 Lexington Avenue, 64th Floor
;              CITY: New York
;              STATE: New York
;              COUNTRY: United States of America
;              ZIP: 10174-6401
;      COMPUTER READABLE FORM:
;              MEDIUM TYPE: Floppy disk
;              COMPUTER: IBM PC compatible
;              OPERATING SYSTEM: PC-DOS/MS-DOS
;              SOFTWARE: PatentIn Release #1.0, Version #1.30
;      CURRENT APPLICATION DATA:
;              APPLICATION NUMBER: US/10/620,651
;              FILING DATE: 16-Jul-2003
;              CLASSIFICATION: <Unknown>
;      PRIOR APPLICATION DATA:
;              APPLICATION NUMBER: 08/932,082
;              FILING DATE: 17-SEPT-1997
;      ATTORNEY/AGENT INFORMATION:
;              NAME: Lambiris, Elias J.
;              REGISTRATION NUMBER: 33,728
;              REFERENCE/DOCKET NUMBER: 4341.204-US
;      TELECOMMUNICATION INFORMATION:
;              TELEPHONE: 212-867-0123
;              TELEFAX: 212-878-9655
;      INFORMATION FOR SEQ ID NO: 19:
;      SEQUENCE CHARACTERISTICS:
;              LENGTH: 138 amino acids
;              TYPE: amino acid
;              TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 19:
US-10-620-651-19

```

```

Query Match          48.4%;  Score 284;  DB 12;  Length 138;
Best Local Similarity 48.2%;  Pred. No. 9.8e-26;
Matches 68;  Conservative 5;  Mismatches 28;  Indels 40;  Gaps 5;

```

```

Qy      2  FPTIPLSRLFDNAMLRAHRLHQLAFDFTYQEFEEAYIPKEQ--KYSFLQ-----N 48
      ||:|      |      :      |      :|      |||      |||      |||      |
Db      3  FPSI-----FTAVLFAASSALAAPVNTTTEDETAQIPAEAVIGYSDLGDFDVAVL PFSN 57

Qy      49  PLGTG-----PRFVNQHLCGSHLVEALYLVCGERGFFYTPK-TR 86
      |      |      |      |      |      |      |      |      |      |
Db      58  STNNGLLFINTTIIASIAAKEEGVSLDKRFVNQHLCGSHLVEALYLVCGERGFFYTPKAAK 117

```

Qy 87 GIVEQCCTSICSLYQLENYCN 107
|||||
Db 118 GIVEQCCTSICSLYQLENYCN 138

RESULT 10

US-10-101-454-33

; Sequence 33, Application US/10101454
; Publication No. US20040110664A1
; GENERAL INFORMATION:
; APPLICANT: Havelund, Svend
; Halstrom, John
; Jonassen, Ib
; Andersen, Asser Sloth
; Markussen, Jan
; TITLE OF INVENTION: ACYLATED INSULIN
; NUMBER OF SEQUENCES: 49
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Novo Nordisk of North America, Inc.
; STREET: 405 Lexington Avenue, 64th Floor
; CITY: New York
; STATE: New York
; COUNTRY: United States of America
; ZIP: 10174-6401
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/10/101,454
; FILING DATE: 20-Mar-2002
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/400,256
; FILING DATE: 03-MAR-1995
; ATTORNEY/AGENT INFORMATION:
; NAME: Lambiris, Elias J.
; REGISTRATION NUMBER: 33,728
; REFERENCE/DOCKET NUMBER: 3985.220-US
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 212-867-0123
; TELEFAX: 212-878-9655
; INFORMATION FOR SEQ ID NO: 33:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 140 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 33:
US-10-101-454-33

Query Match 48.4%; Score 284; DB 16; Length 140;
Best Local Similarity 47.6%; Pred. No. 9.9e-26;
Matches 68; Conservative 6; Mismatches 27; Indels 42; Gaps 5;

Qy 2 FPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQ--KYSFLQ-----N 48

```

Db          ||:|      | : |      :| || || || | || :      |
3 FPSI-----FTAVLF AASSALAAPVNTTTEDETAQIPAEAVIGYSDLEGDFDVAVL PFSN 57

Qy          49 PLGTG-----PRFVNQHLGSHLVEALYLVCGERGFFYTPKT-- 85
          |                                     |||||:
Db          58 STNNGLLFINTTIAAIAAKEEGVSLDKRFVNQHLGSHLVEALYLVCGERGFFYTPKSDD 117

Qy          86 -RGIVEQCCTSICSLYQLENYCN 107
          :|||||
Db          118 AKGIVEQCCTSICSLYQLENYCN 140

```

RESULT 11

US-10-101-454-42

; Sequence 42, Application US/10101454

; Publication No. US20040110664A1

; GENERAL INFORMATION:

; APPLICANT: Havelund, Svend

; Halstrom, John

; Jonassen, Ib

; Andersen, Asser Sloth

; Markussen, Jan

; TITLE OF INVENTION: ACYLATED INSULIN

; NUMBER OF SEQUENCES: 49

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Novo Nordisk of North America, Inc.

; STREET: 405 Lexington Avenue, 64th Floor

; CITY: New York

; STATE: New York

; COUNTRY: United States of America

; ZIP: 10174-6401

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: PatentIn Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/10/101,454

; FILING DATE: 20-Mar-2002

; CLASSIFICATION: <Unknown>

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/400,256

; FILING DATE: 03-MAR-1995

; ATTORNEY/AGENT INFORMATION:

; NAME: Lambiris, Elias J.

; REGISTRATION NUMBER: 33,728

; REFERENCE/DOCKET NUMBER: 3985.220-US

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 212-867-0123

; TELEFAX: 212-878-9655

; INFORMATION FOR SEQ ID NO: 42:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 140 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; SEQUENCE DESCRIPTION: SEQ ID NO: 42:

US-10-101-454-42

Query Match 48.4%; Score 284; DB 16; Length 140;
Best Local Similarity 47.6%; Pred. No. 9.9e-26;
Matches 68; Conservative 6; Mismatches 27; Indels 42; Gaps 5;

```
Qy      2  FPTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQ--KYSFLQ-----N 48
      ||:|      |      :      |      :|      ||      ||      ||      ||      |
Db      3  FPSI-----FTAVLFAASSALAAPVNTTTTETAQIPAEAVIGYSDLEGDFDVAVL PFSN 57

Qy      49  PLGTG-----PRFVNQHLCGSHLVEALYLVCGERGFFYTPKT-- 85
      |      |      |      |      |      |      |      |      |      |      |
Db      58  STNNGLLFINTTIIASIAAKEEGVSMARFVNQHLCGSHLVEALYLVCGERGFFYTPKSDD 117

Qy      86  -RGIVEQCCTSICSLYQLENYCN 107
      :|||||||||||||||||
Db     118  AKGIVEQCCTSICSLYQLENYCN 140
```

RESULT 12

US-10-101-454-15

```
; Sequence 15, Application US/10101454
; Publication No. US20040110664A1
;   GENERAL INFORMATION:
;       APPLICANT: Havelund, Svend
;                   Halstrom, John
;                   Jonassen, Ib
;                   Andersen, Asser Sloth
;                   Markussen, Jan
;   TITLE OF INVENTION: ACYLATED INSULIN
;   NUMBER OF SEQUENCES: 49
;   CORRESPONDENCE ADDRESS:
;       ADDRESSEE: Novo Nordisk of North America, Inc.
;       STREET: 405 Lexington Avenue, 64th Floor
;       CITY: New York
;       STATE: New York
;       COUNTRY: United States of America
;       ZIP: 10174-6401
;   COMPUTER READABLE FORM:
;       MEDIUM TYPE: Floppy disk
;       COMPUTER: IBM PC compatible
;       OPERATING SYSTEM: PC-DOS/MS-DOS
;       SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;       APPLICATION NUMBER: US/10/101,454
;       FILING DATE: 20-Mar-2002
;       CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;       APPLICATION NUMBER: 08/400,256
;       FILING DATE: 03-MAR-1995
;   ATTORNEY/AGENT INFORMATION:
;       NAME: Lambiris, Elias J.
;       REGISTRATION NUMBER: 33,728
;       REFERENCE/DOCKET NUMBER: 3985.220-US
;   TELECOMMUNICATION INFORMATION:
;       TELEPHONE: 212-867-0123
;       TELEFAX: 212-878-9655
```



```
; INFORMATION FOR SEQ ID NO: 15:
; SEQUENCE CHARACTERISTICS:
;     LENGTH: 104 amino acids
;     TYPE: amino acid
;     TOPOLOGY: linear
;     MOLECULE TYPE: protein
;     SEQUENCE DESCRIPTION: SEQ ID NO: 15:
US-10-101-454-15
```

```
Query Match          47.9%;  Score 281;  DB 16;  Length 104;
Best Local Similarity 71.8%;  Pred. No. 1.6e-25;
Matches   56;  Conservative    6;  Mismatches    8;  Indels    8;  Gaps    3;
```

```
Qy      37 IPKEQ---KYSFLQNPLGTGPRFVNQHLCGSHLVEALYLVCGERGFFYTPKT---RGIV 89
      ||:| : : | | : ||||| : : |||
Db      28 IPEESLIIAENTTLAN-VAMAKRFVNQHLCGSHLVEALYLVCGERGFFYTPKSDDAKGIV 86

Qy      90 EQCCTSICSLYQLENYCN 107
      |||||
Db      87 EQCCTSICSLYQLENYCN 104
```

RESULT 13

```
US-09-858-935B-5
; Sequence 5, Application US/09858935B
; Publication No. US20030069177A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Filvaroff, Ellen
; APPLICANT: Lowman, Henry B.
; TITLE OF INVENTION: METHOD FOR TREATING CARTILAGE DISORDERS
; FILE REFERENCE: P1794R1
; CURRENT APPLICATION NUMBER: US/09/858,935B
; CURRENT FILING DATE: 2002-07-02
; PRIOR APPLICATION NUMBER: US 60/248,985
; PRIOR FILING DATE: 2000-11-15
; PRIOR APPLICATION NUMBER: US 60/204,490
; PRIOR FILING DATE: 2000-05-16
; NUMBER OF SEQ ID NOS: 153
; SEQ ID NO 5
;     LENGTH: 51
;     TYPE: PRT
;     ORGANISM: Homo sapiens
US-09-858-935B-5
```

```
Query Match          47.4%;  Score 278.5;  DB 10;  Length 51;
Best Local Similarity 98.1%;  Pred. No. 1.3e-25;
Matches   51;  Conservative    0;  Mismatches    0;  Indels    1;  Gaps    1;
```

```
Qy      56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
      |||||
Db      1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51
```

RESULT 14

```
US-10-444-649-3
; Sequence 3, Application US/10444649
```

; Publication No. US20040033951A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,649
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US/09/724,479
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/477,923
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-444-649-3

Query Match 47.4%; Score 278.5; DB 12; Length 51;
Best Local Similarity 98.1%; Pred. No. 1.3e-25;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|||||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

RESULT 15
US-10-444-701-3
; Sequence 3, Application US/10444701
; Publication No. US20040033952A1
; GENERAL INFORMATION:
; APPLICANT: Dubaquie, Yves
; APPLICANT: Lowman, Henry
; TITLE OF INVENTION: PROTEIN VARIANTS
; FILE REFERENCE: P1712R1
; CURRENT APPLICATION NUMBER: US/10/444,701
; CURRENT FILING DATE: 2003-05-22
; PRIOR APPLICATION NUMBER: US/09/723,866
; PRIOR FILING DATE: 2000-11-28
; PRIOR APPLICATION NUMBER: US/09/477,923
; PRIOR FILING DATE: 2000-01-05
; NUMBER OF SEQ ID NOS: 6
; SEQ ID NO 3
; LENGTH: 51
; TYPE: PRT
; ORGANISM: Homo sapiens
US-10-444-701-3

Query Match 47.4%; Score 278.5; DB 12; Length 51;
Best Local Similarity 98.1%; Pred. No. 1.3e-25;
Matches 51; Conservative 0; Mismatches 0; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
|||||
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTSICSLYQLENYCN 51

Search completed: July 15, 2004, 17:05:09
Job time : 45.5168 secs

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:29:50 ; Search time 36.5317 Seconds
(without alignments)
924.141 Million cell updates/sec

Title: US-09-423-100-6
Perfect score: 587
Sequence: 1 MFPTIPLSRLFDNAMLRAHR.....IVEQCCTSICSPLYQLENYCN 107

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 1017041 seqs, 315518202 residues

Total number of hits satisfying chosen parameters: 1017041

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SPTREMBL_25:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp Vertebrate:*
14: sp_unclassified:*
15: sp_rvirus:*
16: sp_bacteriap:*
17: sp_archeap:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result	%	Query					
No.	Score	Match	Length	DB	ID		Description

1	267	45.5	110	6	Q8HXV2	Q8h xv2 pongo pygma
2	251	42.8	110	6	Q8WNW6	Q8wnw6 felis silve
3	249	42.4	217	6	Q8WNE0	Q8wne0 ateles geof
4	228	38.8	245	4	O14644	O14644 homo sapien
5	219.5	37.4	106	13	Q9I8Q7	Q9i8q7 rana pipien
6	215	36.6	184	6	Q866T9	Q866t9 pan troglod
7	213	36.3	212	6	Q07368	Q07368 macaca mula
8	213	36.3	217	6	Q07367	Q07367 macaca mula
9	205	34.9	217	6	Q866U1	Q866u1 pan troglod
10	201.5	34.3	110	13	Q98TA8	Q98ta8 pantodon bu
11	201	34.2	217	6	Q07369	Q07369 macaca mula
12	201	34.2	217	6	Q866T8	Q866t8 pan troglod
13	197	33.6	217	4	Q14407	Q14407 homo sapien
14	195.5	33.3	108	13	Q9DDE5	Q9dde5 brachydanio
15	195.5	33.3	108	13	Q90ZN4	Q90zn4 catla catla
16	195	33.2	111	13	Q98TB0	Q98tb0 chitala chi
17	195	33.2	217	6	Q8WND9	Q8wnd9 ateles geof
18	193.5	33.0	110	13	Q90ZY1	Q90zy1 hiodon alos
19	191.5	32.6	111	13	Q98TA7	Q98ta7 osteoglossu
20	189.5	32.3	108	13	Q98TB1	Q98tb1 catostomus
21	187.5	31.9	87	13	Q98TA9	Q98ta9 gnathonemus
22	186	31.7	217	6	Q866U0	Q866u0 pan troglod
23	185.5	31.6	91	13	Q98TB2	Q98tb2 ambloplites
24	171.5	29.2	217	6	Q8MI74	Q8mi74 callithrix
25	161	27.4	216	11	O70615	O70615 spalax leuc
26	159.5	27.2	52	6	Q9TV91	Q9tv91 equus cabal
27	159.5	27.2	216	6	Q8MI73	Q8mi73 delphinus d
28	159.5	27.2	216	6	Q8HYE5	Q8h ye5 ailuropoda
29	159.5	27.2	216	6	Q7YQB8	Q7yqb8 hippopotamu
30	156	26.6	216	11	Q9R2C3	Q9r2c3 mus musculu
31	155.5	26.5	216	6	Q7YRR6	Q7yrr6 camelus dro
32	154	26.2	216	11	Q9JKM4	Q9jkm4 cavia porce
33	152	25.9	178	6	Q95MJ5	Q95mj5 tarsius ban
34	149.5	25.5	204	6	Q95205	Q95205 ovis aries
35	149	25.4	132	13	Q8AV14	Q8av14 petromyzon
36	148.5	25.3	104	13	Q7T107	Q7t107 dicentrarch
37	148.5	25.3	108	13	Q800N0	Q800n0 morone chry
38	148.5	25.3	108	13	Q800M9	Q800m9 morone saxa
39	148.5	25.3	108	13	Q800M8	Q800m8 morone chry
40	148.5	25.3	108	13	Q800M7	Q800m7 morone amer
41	148.5	25.3	159	13	O93607	O93607 paralichthy
42	148.5	25.3	182	13	O73720	O73720 oreochromis
43	148.5	25.3	182	13	O42289	O42289 oreochromis
44	148.5	25.3	182	13	P79824	P79824 oreochromis
45	148.5	25.3	185	13	O57436	O57436 paralichthy

ALIGNMENTS

RESULT 1

Q8HXV2

ID Q8HXV2 PRELIMINARY; PRT; 110 AA.

AC Q8HXV2;

DT 01-MAR-2003 (TrEMBLrel. 23, Created)

DT 01-MAR-2003 (TrEMBLrel. 23, Last sequence update)

DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)

DE Insulin precursor.
 GN INS.
 OS Pongo pygmaeus (Orangutan).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pongo.
 OX NCBI_TaxID=9600;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Stead J.D.H., Jeffreys A.J.;
 RT "Haplotype diversity at the insulin region."
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY137503; AAN06937.1; -.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 SQ SEQUENCE 110 AA; 12038 MW; 22D2B32B94F520F8 CRC64;

Query Match 45.5%; Score 267; DB 6; Length 110;
 Best Local Similarity 60.5%; Pred. No. 9e-24;
 Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

Qy 56 FVNQHLGSHLVEALYLVCGERGFFYTPKT----- 85
 ||||||||||||||||||||||||||||
 Db 25 FVNQHLGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84
 Qy 86 ----RGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||
 Db 85 SLQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 2

Q8WNW6

ID Q8WNW6 PRELIMINARY; PRT; 110 AA.
 AC Q8WNW6;
 DT 01-MAR-2002 (TrEMBLrel. 20, Created)
 DT 01-MAR-2002 (TrEMBLrel. 20, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Preproinsulin.
 OS Felis silvestris catus (Cat).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
 OX NCBI_TaxID=9685;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RA Okamoto S., Morimatsu M.;
 RT "cat insulin."
 RL Submitted (MAY-2000) to the EMBL/GenBank/DDBJ databases.
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AB043535; BAB84110.1; -.
 DR GO; GO:0005576; C:extracellular; IEA.

OX NCBI_TaxID=8404;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=20362507; PubMed=10818274;
 RA Irwin D.M., Sivarajah P.;
 RT "Proinsulin cDNAs from the leopard frog, *Rana pipiens*: evolution of
 RT proinsulin processing.";
 RL Comp. Biochem. Physiol. 125B:405-410(2000).
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AF227187; AAF87285.1; -.
 DR HSSP; P01315; 1SDB.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 SQ SEQUENCE 106 AA; 12183 MW; 3A870EEC70217F92 CRC64;

Query Match 37.4%; Score 219.5; DB 13; Length 106;
 Best Local Similarity 49.4%; Pred. No. 3.7e-18;
 Matches 41; Conservative 7; Mismatches 4; Indels 31; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTR----- 86
 | ||:|||||||:||||:|::|
 Db 24 FDNQYLCGSHLVEALYMCVCGDRGFFYSPRSRRDLEQPLVNGLQGSSELDQMVSQAFQKR 83

 Qy 87 --GIVEQCCTSI CSLYQLENYCN 107
 ||||| : |||| |||||
 Db 84 KPGIVEQCCHNTCSLYDLENYCN 106

RESULT 6

Q866T9

ID Q866T9 PRELIMINARY; PRT; 184 AA.
 AC Q866T9;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Placental lactogen PL-C (Fragment).
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Revol A., Esquivel D.E., Barrera H.S.;
 RT "The GH-PL locus a hot-point between human and chimpanzee genomes.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY146627; AAN84507.1; -.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR InterPro; IPR001400; Somatotropin.
 DR Pfam; PF00103; hormone; 1.

RESULT 8

Q07367

ID Q07367 PRELIMINARY; PRT; 217 AA.
 AC Q07367;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Chorionic somatomammotropin-1.
 OS Macaca mulatta (Rhesus macaque).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
 OC Cercopithecinae; Macaca.
 OX NCBI_TaxID=9544;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Midpregnancy placenta;
 RX MEDLINE=94008724; PubMed=8404617;
 RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;
 RT "Cloning of four growth hormone/chorionic somatomammotropin-related
 RT complementary deoxyribonucleic acids differentially expressed during
 RT pregnancy in the rhesus monkey placenta.";
 RL Endocrinology 133:1744-1752(1993).
 DR EMBL; L16552; AAA18839.1; -.
 DR PIR; I53267; I53267.
 DR HSSP; P01241; LAXI.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR InterPro; IPR001400; Somatotropin.
 DR Pfam; PF00103; hormone; 1.
 DR PRINTS; PR00836; SOMATOTROPIN.
 DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
 SQ SEQUENCE 217 AA; 24942 MW; FF5AA8915131F2BC CRC64;

Query Match 36.3%; Score 213; DB 6; Length 217;
 Best Local Similarity 78.7%; Pred. No. 5e-17;
 Matches 37; Conservative 9; Mismatches 1; Indels 0; Gaps 0;

Qy 3 PTIPLSRLFDNAMLRAHRLHQLAFDITYQEFEEAYIPKEQKYSFLQNP 49
 |:|||||||:|:|||||||||||||||||||||:|:| :||
 Db 28 PSVPLSRLFDHAMIQAHLHQLAFDITYQEFEEAYIPKEKKHSLMENP 74

RESULT 9

Q866U1

ID Q866U1 PRELIMINARY; PRT; 217 AA.
 AC Q866U1;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Placental lactogen PL-A.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Revol A., Esquivel D.E., Barrera H.S.;

RT "The GH-PL locus a hot-point between human and chimpanzee genomes.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY146625; AAN84505.1; -.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR InterPro; IPR001400; Somatotropin.
 DR Pfam; PF00103; hormone; 1.
 DR PRINTS; PR00836; SOMATOTROPIN.
 DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
 DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
 SQ SEQUENCE 217 AA; 25081 MW; C74B6262D8A93060 CRC64;

Query Match 34.9%; Score 205; DB 6; Length 217;
 Best Local Similarity 84.8%; Pred. No. 4.4e-16;
 Matches 39; Conservative 4; Mismatches 3; Indels 0; Gaps 0;

Qy 4 TIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49
 |:|||||||:||||:||||:||||| ||||| ||||| ||||| |||||
 Db 29 TVPLSRLFDHAMLQAHRAYQLAIDTYQEFEEAYILKEQKYSFLQNP 74

RESULT 10

Q98TA8

ID Q98TA8 PRELIMINARY; PRT; 110 AA.
 AC Q98TA8;
 DT 01-JUN-2001 (TrEMBLrel. 17, Created)
 DT 01-JUN-2001 (TrEMBLrel. 17, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Preproinsulin.
 OS Pantodon buchholtzi (Butterflyfish).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Osteoglossomorpha;
 OC Osteoglossiformes; Pantodontidae; Pantodon.
 OX NCBI_TaxID=8276;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21203577; PubMed=11306171;
 RA Al-Mahrouki A.A., Irwin D.M., Graham L.C., Youson J.H.;
 RT "Molecular cloning of preproinsulin cDNAs from several
 RT osteoglossomorphs and a cyprinid."
 RL Mol. Cell. Endocrinol. 174:51-58(2001).
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AF199588; AAK28712.1; -.
 DR HSSP; P01308; 1HIS.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 SQ SEQUENCE 110 AA; 12324 MW; BDECCD659D872E06 CRC64;

Query Match 34.3%; Score 201.5; DB 13; Length 110;
 Best Local Similarity 43.5%; Pred. No. 5.2e-16;

Matches 37; Conservative 8; Mismatches 5; Indels 35; Gaps 1;

```
Qy      58 NQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
          :|||||||:|:|:|:|:|:|:|
Db      26 SQHLCGSHLVDALYMVCGEKGFFYQPKTKRDVDPLLGLFLSPKSAQENEADEYPYKDQGD 85

Qy      86 ---RGIVEQCCTSICSLYQLENYCN 107
          ||||| |::|:|
Db      86 KVKRGIVEQCCHHPCNIFDLQNYCN 110
```

RESULT 11

Q07369

ID Q07369 PRELIMINARY; PRT; 217 AA.
AC Q07369;
DT 01-NOV-1996 (TrEMBLrel. 01, Created)
DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
DE Chorionic somatomammotropin-3.
OS Macaca mulatta (Rhesus macaque).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9544;
RN [1]
RP SEQUENCE FROM N.A.
RC TISSUE=Midpregnancy placenta;
RX MEDLINE=94008724; PubMed=8404617;
RA Golos T.G., Durning M., Fisher J.M., Fowler P.D.;
RT "Cloning of four growth hormone/chorionic somatomammotropin-related
RT complementary deoxyribonucleic acids differentially expressed during
RT pregnancy in the rhesus monkey placenta.";
RL Endocrinology 133:1744-1752(1993).
DR EMBL; L16554; AAA18841.1; -.
DR PIR; I67409; I67409.
DR HSSP; P01241; IAXI.
DR GO; GO:0005576; C:extracellular; IEA.
DR GO; GO:0005179; F:hormone activity; IEA.
DR InterPro; IPR001400; Somatotropin.
DR Pfam; PF00103; hormone; 1.
DR PRINTS; PR00836; SOMATOTROPIN.
DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
SQ SEQUENCE 217 AA; 24874 MW; F1EB6AFDBBA1B185 CRC64;

Query Match 34.2%; Score 201; DB 6; Length 217;
Best Local Similarity 74.5%; Pred. No. 1.3e-15;
Matches 35; Conservative 8; Mismatches 4; Indels 0; Gaps 0;

```
Qy      3 PTIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQNP 49
          |::||||| |::|||||||:|:|:|:|:|
Db      28 PSVPLSRLFDNIMQAHLRLHQLAFDTYQEFEKTYIPKEKKHSLMGNP 74
```

RESULT 12

Q866T8

ID Q866T8 PRELIMINARY; PRT; 217 AA.

AC Q866T8;
 DT 01-JUN-2003 (TrEMBLrel. 24, Created)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Placental lactogen PL-D.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Revol A., Esquivel D.E., Barrera H.S.;
 RT "The GH-PL locus a hot-point between human and chimpanzee genomes.";
 RL Submitted (AUG-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; AY146628; AAN84508.1; -.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR InterPro; IPR001400; Somatotropin.
 DR Pfam; PF00103; hormone; 1.
 DR PRINTS; PR00836; SOMATOTROPIN.
 DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
 DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
 SQ SEQUENCE 217 AA; 25135 MW; 1EB7B89B8A12E4F4 CRC64;

Query Match 34.2%; Score 201; DB 6; Length 217;
 Best Local Similarity 82.2%; Pred. No. 1.3e-15;
 Matches 37; Conservative 5; Mismatches 3; Indels 0; Gaps 0;

Qy 4 TIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQN 48
 |:|||||:||||:|||| |||| |||||:||||| :
 Db 29 TVPLSRLFDHAMLQAHRAHQLAIDTYQEFEEAYIPKDQKYSFLHD 73

RESULT 13

Q14407

ID Q14407 PRELIMINARY; PRT; 217 AA.
 AC Q14407;
 DT 01-NOV-1996 (TrEMBLrel. 01, Created)
 DT 01-NOV-1996 (TrEMBLrel. 01, Last sequence update)
 DT 01-OCT-2003 (TrEMBLrel. 25, Last annotation update)
 DE Chorionic somatomammotropin CS-2 (Chorionic somatomammotropin hormone 2).
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 OX NCBI_TaxID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=89307277; PubMed=2744760;
 RA Chen E.Y., Liao Y.C., Smith D.H., Barrera-Saldana H.A., Gelinas R.E.,
 RA Seeburg P.H.;
 RT "The human growth hormone locus: nucleotide sequence, biology, and
 RT evolution.";
 RL Genomics 4:479-497(1989).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=91102558; PubMed=1980158;

RA Vnencak-Jones C.L., Phillips J.A. III.;
 RT "Hot spots for growth hormone gene deletions in homologous regions
 RT outside of Alu repeats.";
 RL Science 250:1745-1748(1990).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=Placenta;
 RA Strausberg R.;
 RL Submitted (JUL-2002) to the EMBL/GenBank/DDBJ databases.
 DR EMBL; J03071; AAA52553.1; -.
 DR EMBL; BC022044; AAH22044.1; -.
 DR EMBL; BC035965; AAH35965.1; -.
 DR PIR; E32435; E32435.
 DR HSSP; P01241; 1A22.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR InterPro; IPR001400; Somatotropin.
 DR Pfam; PF00103; hormone; 1.
 DR PRINTS; PR00836; SOMATOTROPIN.
 DR PROSITE; PS00266; SOMATOTROPIN_1; 1.
 DR PROSITE; PS00338; SOMATOTROPIN_2; 1.
 SQ SEQUENCE 217 AA; 24994 MW; 39FAACDDB6B2E951 CRC64;

Query Match 33.6%; Score 197; DB 4; Length 217;
 Best Local Similarity 80.0%; Pred. No. 3.9e-15;
 Matches 36; Conservative 5; Mismatches 4; Indels 0; Gaps 0;

QY 4 TIPLSRLFDNAMLRAHRLHQLAFDTYQEFEEAYIPKEQKYSFLQN 48
 |:|||||:|:|:| | | | | | | | | | | | | | | | :
 Db 29 TVPLSRLFDHAMLQAHRAHQLAIDTYQEFETYIPKDQKYSFLHD 73

RESULT 14

Q9DDE5

ID Q9DDE5 PRELIMINARY; PRT; 108 AA.
 AC Q9DDE5;
 DT 01-MAR-2001 (TrEMBLrel. 16, Created)
 DT 01-MAR-2001 (TrEMBLrel. 16, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Brachydanio rerio (Zebrafish) (Danio rerio).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Danio.
 OX NCBI_TaxID=7955;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=99425190; PubMed=10495291;
 RA Argenton F., Zecchin E., Bortolussi M.;
 RT "Early appearance of pancreatic hormone-expressing cells in the
 RT zebrafish embryo.";
 RL Mech. Dev. 87:217-221(1999).
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AJ237750; CAC20109.1; -.
 DR HSSP; P01308; 1LPH.

DR ZFIN; ZDB-GENE-980526-110; ins.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Signal.
 FT SIGNAL 1 23 POTENTIAL.
 FT CHAIN 24 53 INSULIN B CHAIN.
 FT CHAIN 86 108 INSULIN A CHAIN.
 SQ SEQUENCE 108 AA; 11904 MW; 3195289E72AD6D25 CRC64;

Query Match 33.3%; Score 195.5; DB 13; Length 108;
 Best Local Similarity 45.1%; Pred. No. 2.6e-15;
 Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

Qy 59 QHLCGSHLVEALYLVCGERGFFYTPK-----T 85
 |||||:||||| ||| ||
 Db 27 QHLCGSHLVDALYLVCGPTGFFYNPKRDVEPLLGLFPKSAQETEVADFAFKDHAELIRK 86
 Qy 86 RGIVEQCCTSICSLYQLENYCN 107
 ||||| ||::|:||||
 Db 87 RGIVEQCCHKPCSIFELQNYCN 108

RESULT 15

Q90ZN4

ID Q90ZN4 PRELIMINARY; PRT; 108 AA.
 AC Q90ZN4;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2003 (TrEMBLrel. 24, Last annotation update)
 DE Preproinsulin.
 OS Catla catla (catla).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Actinopterygii; Neopterygii; Teleostei; Ostariophysi; Cypriniformes;
 OC Cyprinidae; Catla.
 OX NCBI_TaxID=72446;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bhattacharya S., Roy S.S., Dasgupta S., Ravikumar L., Mukherjee M.,
 RA Bandyopadhyaya I., Wakabayashi K.;
 RT "A new cell secreting insulin."
 RL Submitted (APR-2001) to the EMBL/GenBank/DDBJ databases.
 CC -!- SUBCELLULAR LOCATION: SECRETED (BY SIMILARITY).
 CC -!- SIMILARITY: BELONGS TO THE INSULIN/IGF/RELAXIN FAMILY.
 DR EMBL; AF373021; AAK51558.1; -.
 DR HSSP; P01308; 1LNP.
 DR GO; GO:0005576; C:extracellular; IEA.
 DR GO; GO:0005179; F:hormone activity; IEA.
 DR GO; GO:0007582; P:physiological processes; IEA.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.

DR SMART; SM00078; ILGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
SQ SEQUENCE 108 AA; 11881 MW; D713026E22EF5D59 CRC64;

Query Match 33.3%; Score 195.5; DB 13; Length 108;
Best Local Similarity 45.1%; Pred. No. 2.6e-15;
Matches 37; Conservative 5; Mismatches 7; Indels 33; Gaps 1;

Qy 59 QHLCGSHLVEALYLVCGERGFFYTPK-----T 85
|||||||:||||| ||| ||
Db 27 QHLCGSHLVDALYLVCGPTGFFYNPKRDVDPLMGFLPPKSAQETEVADFAFKDHAEVIRK 86

Qy 86 RGIVEQCCTSICSLYQLENYCN 107
||||||| ||::|:||||
Db 87 RGIVEQCCHKPCSIFELQNYCN 108

Search completed: July 15, 2004, 16:41:00
Job time : 36.6984 secs

OM protein - protein search, using sw model

Run on: July 15, 2004, 16:28:49 ; Search time 7.38619 Seconds
 (without alignments)
 754.314 Million cell updates/sec

Title: US-09-423-100-6
 Perfect score: 587
 Sequence: 1 MFPTIPLSRLFDNAMLRAHR.....IVEQCCTSICSLYQLENYCN 107

Scoring table: BLOSUM62
 Gapop 10.0 , Gapext 0.5

Searched: 141681 seqs, 52070155 residues

Total number of hits satisfying chosen parameters: 141681

Minimum DB seq length: 0
 Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
 Maximum Match 100%
 Listing first 45 summaries

Database : SwissProt_42:*

Pred. No. is the number of results predicted by chance to have a
 score greater than or equal to the score of the result being printed,
 and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Query		DB	ID	Description
		Match	Length			
1	273.5	46.6	51	1	INS_BALPH	P01312 balaenopter
2	273.5	46.6	51	1	INS_ELEMA	P01316 elephas max
3	273	46.5	110	1	INS_CERAE	P30407 cercopithec
4	273	46.5	110	1	INS_MACFA	P30406 macaca fasc
5	268.5	45.7	51	1	INS_ACOCA	P01324 acomys cahi
6	267	45.5	110	1	INS_HUMAN	P01308 homo sapien
7	267	45.5	110	1	INS_PANTR	P30410 pan troglod
8	266	45.3	110	1	INS_SPETR	Q91xi3 spermophilu
9	263.5	44.9	51	1	INS_BALBO	P01314 balaenopter
10	263.5	44.9	51	1	INS_CAMDR	P01320 camelus dro
11	263.5	44.9	51	1	INS_CAPHI	P01319 capra hircu
12	263	44.8	108	1	INS_PIG	P01315 sus scrofa
13	263	44.8	110	1	INS_RABIT	P01311 oryctolagus
14	262.5	44.7	51	1	INS_FELCA	P06306 felis silve
15	262	44.6	110	1	INS_CANFA	P01321 canis famil
16	260	44.3	110	1	INS_CRILO	P01313 cricetulus
17	258.5	44.0	105	1	INS_BOVIN	P01317 bos taurus

18	257	43.8	108	1	INS_AOTTR	P10604	aotus trivi
19	257	43.8	110	1	INS_PSAOB	Q62587	psammomys o
20	256.5	43.7	51	1	INS_DIDMA	P18109	didelphis m
21	255.5	43.5	217	1	SOMA_HUMAN	P01241	homo sapien
22	255.5	43.5	217	1	SOMA_MACMU	P33093	macaca mula
23	255.5	43.5	217	1	SOMA_PANTR	P58756	pan troglod
24	254.5	43.4	105	1	INS_SHEEP	P01318	ovis aries
25	252	42.9	86	1	INS_HORSE	P01310	equus cabal
26	251.5	42.8	51	1	INS_CHIBR	P01327	chinchilla
27	250	42.6	108	1	INS1_MOUSE	P01325	mus musculu
28	249	42.4	110	1	INS1_RAT	P01322	rattus norv
29	249	42.4	217	1	SOMA_CALJA	Q9gmb3	callithrix
30	249	42.4	217	1	SOMA_SAIIBB	P58343	saimiri bol
31	248.5	42.3	51	1	INS_ANSAN	P07454	anser anser
32	248	42.2	110	1	INS2_MOUSE	P01326	mus musculu
33	248	42.2	110	1	INS2_RAT	P01323	rattus norv
34	246	41.9	52	1	INS_ACIGU	P81423	acipenser g
35	245	41.7	103	1	INS_SELRF	P51463	selasphorus
36	244.5	41.7	51	1	INS_HYSCR	P01328	hystrix cri
37	244.5	41.7	51	1	INS_TRASC	P31887	trachemys s
38	239.5	40.8	107	1	INS_CHICK	P01332	gallus gall
39	236	40.2	217	1	SOM2_PANTR	P58757	pan troglod
40	235.5	40.1	51	1	INS_ORNAN	Q9tqy7	ornithorhyn
41	233.5	39.8	81	1	INS_ANAPL	P01333	anas platyr
42	231.5	39.4	51	1	INS_ALLMI	P12703	alligator m
43	231	39.4	52	1	INS_LEPSP	P09476	lepisosteus
44	228.5	38.9	51	1	INS_ZAODH	P12708	zaocys dhum
45	228	38.8	217	1	SOM2_HUMAN	P01242	homo sapien

ALIGNMENTS

RESULT 1

INS_BALPH

ID INS_BALPH STANDARD; PRT; 51 AA.

AC P01312;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin.

GN INS.

OS Balaenoptera physalus (Finback whale) (Common rorqual), and

OS Physeter catodon (Sperm whale) (Physeter macrocephalus).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;

OC Balaenopteridae; Balaenoptera.

OX NCBI_TaxID=9770, 9755;

RN [1]

RP PARTIAL SEQUENCE.

RC SPECIES=B.physalus;

RA Hama H., Titani K., Sakaki S., Narita K.;

RT "The amino acid sequence in fin-whale insulin.";

RL J. Biochem. 56:285-293(1964).

RN [2]

RP SEQUENCE.

RC SPECIES=P.catodon;

RA Ishihara Y., Saito T., Ito Y., Fujino M.;
 RT "Structure of sperm- and sei-whale insulins and their breakdown by
 RT whale pepsin.";
 RL Nature 181:1468-1469(1958).
 RN [3]
 RP SEQUENCE.
 RC SPECIES=P.catodon;
 RA Harris J.I., Sanger F., Naughton M.A.;
 RT "Species differences in insulin.";
 RL Arch. Biochem. Biophys. 65:427-438(1956).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR PIR; A91918; INWHF.
 DR PIR; A93142; INWHP.
 DR HSSP; P01317; 1APH.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA; 5766 MW; 9007B514691A7CDD CRC64;

Query Match 46.6%; Score 273.5; DB 1; Length 51;
 Best Local Similarity 96.2%; Pred. No. 1.9e-22;
 Matches 50; Conservative 0; Mismatches 1; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIICSLYQLENYCN 107
 |||||
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCTSIICSLYQLENYCN 51

RESULT 2
 INS_ELEMA
 ID INS_ELEMA STANDARD; PRT; 51 AA.
 AC P01316;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin.
 GN INS.
 OS Elephas maximus (Indian elephant).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Proboscidea; Elephantidae; Elephas.
 OX NCBI_TaxID=9783;
 RN [1]

RP SEQUENCE.
 RX MEDLINE=66160119; PubMed=5949593;
 RA Smith L.F.;
 RT "Species variation in the amino acid sequence of insulin.";
 RL Am. J. Med. 40:662-666(1966).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- MISCELLANEOUS: THE SPECIES OF ELEPHANT IS NOT GIVEN, BUT IT IS
 CC MOST PROBABLY THE INDIAN ELEPHANT (ELEPHAS MAXIMUS).
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR HSSP; P01308; 1AI0.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA; 5752 MW; 9007B50CDB457D6D CRC64;

Query Match 46.6%; Score 273.5; DB 1; Length 51;
 Best Local Similarity 94.2%; Pred. No. 1.9e-22;
 Matches 49; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 ||||| :|||||
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKT-GIVEQCCTGVCSLYQLENYCN 51

RESULT 3
 INS_CERA
 ID INS_CERA STANDARD; PRT; 110 AA.
 AC P30407; P01309;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Cercopithecus aethiops (Green monkey) (Grivet).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
 OC Cercopithecinae; Cercopithecus.
 OX NCBI_TaxID=9534;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219953; PubMed=1560757;
 RA Seino S., Bell G.I., Li W.;
 RT "Sequences of primate insulin genes support the hypothesis of a

RESULT 4

INS_MACFA

ID INS_MACFA STANDARD; PRT; 110 AA.
AC P30406; P01309;
DT 21-JUL-1986 (Rel. 01, Created)
DT 13-AUG-1987 (Rel. 05, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin precursor.
GN INS.
OS Macaca fascicularis (Crab eating macaque) (Cynomolgus monkey).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;
OC Cercopithecinae; Macaca.
OX NCBI_TaxID=9541;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=83080474; PubMed=6184262;
RA Wetekam W., Groneberg J., Leineweber M., Wengenmayer F.,
RA Winnacker E.-L.;
RT "The nucleotide sequence of cDNA coding for preproinsulin from the
RT primate Macaca fascicularis.";
RL Gene 19:179-183(1982).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; J00336; AAA36849.1; -.
DR PIR; JQ0178; JQ0178.
DR HSSP; P01308; 1AI0.
DR InterPro; IPR004825; Ins/IGF/relax.
DR Pfam; PF00049; Insulin; 1.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism; Signal.
FT SIGNAL 1 24
FT CHAIN 25 54 INSULIN B CHAIN.
FT PROPEP 57 87 C PEPTIDE.
FT CHAIN 90 110 INSULIN A CHAIN.
FT DISULFID 31 96 INTERCHAIN.
FT DISULFID 43 109 INTERCHAIN.
FT DISULFID 95 100

SQ SEQUENCE 110 AA; 11991 MW; 83C6E33A80A420F9 CRC64;

Query Match 46.5%; Score 273; DB 1; Length 110;
Best Local Similarity 60.2%; Pred. No. 4.9e-22;
Matches 53; Conservative 0; Mismatches 1; Indels 34; Gaps 1;

```
Qy      54 PRFVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
          | | | | | | | | | | | | | | | | | | | | | |
Db      23 PAFVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDPQVGQVELGGGPGAGSLQPLAL 82

Qy      86 -----RGIVEQCCTSICSLYQLENYCN 107
          | | | | | | | | | | | | | | | | | | | | | |
Db      83 EGS LQKRGIVEQCCTSICSLYQLENYCN 110
```

RESULT 5

INS_ACOCA

ID INS_ACOCA STANDARD; PRT; 51 AA.

AC P01324;

DT 21-JUL-1986 (Rel. 01, Created)

DT 21-JUL-1986 (Rel. 01, Last sequence update)

DT 10-OCT-2003 (Rel. 42, Last annotation update)

DE Insulin.

GN INS.

OS *Acomys cahirinus* (Egyptian spiny mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; *Acomys*.

OX NCBI_TaxID=10068;

RN [1]

RP COMPOSITION.

RX MEDLINE=72189454; PubMed=5028210;

RA Buenzli H.F., Humbel R.E.;

RT "Isolation and partial structural analysis of insulin from mouse (*Mus musculus*) and spiny mouse (*Acomys cahirinus*).";

RL Hoppe-Seyler's Z. Physiol. Chem. 353:444-450(1972).

CC -!- FUNCTION: Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver.

CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two disulfide bonds.

CC -!- SUBCELLULAR LOCATION: Secreted.

CC -!- SIMILARITY: Belongs to the insulin family.

DR PIR; A01591; INMSSP.

DR HSSP; P01308; 1TYM.

DR InterPro; IPR004825; Ins/IGF/relax.

DR PRINTS; PR00277; INSULINB.

DR SMART; SM00078; IIGF; 1.

DR PROSITE; PS00262; INSULIN; 1.

KW Insulin family; Hormone; Glucose metabolism.

FT CHAIN 1 30 INSULIN B CHAIN.

FT NON_CONS 30 31

FT CHAIN 31 51 INSULIN A CHAIN.

FT DISULFID 7 37 INTERCHAIN (BY SIMILARITY).

FT DISULFID 19 50 INTERCHAIN (BY SIMILARITY).

FT DISULFID 36 41 BY SIMILARITY.

SQ SEQUENCE 51 AA; 5768 MW; 992BD8B629047D3D CRC64;

Query Match 45.7%; Score 268.5; DB 1; Length 51;
Best Local Similarity 92.3%; Pred. No. 6.6e-22;
Matches 48; Conservative 3; Mismatches 0; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
||:|||||||||||||||||||||||: ||:|||||||||||||||
Db 1 FVBQHLCGSHLVEALYLVCGERGFFYTPKS-GIVDQCCTSICSLYQLENYCN 51

RESULT 6

INS_HUMAN

ID INS_HUMAN STANDARD; PRT; 110 AA.
AC P01308;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 15-MAR-2004 (Rel. 43, Last annotation update)
DE Insulin precursor.
GN INS.
OS Homo sapiens (Human).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
OX NCBI_TaxID=9606;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=80120725; PubMed=6243748;
RA Bell G.I., Pictet R.L., Rutter W.J., Cordell B., Tischler E.,
RA Goodman H.M.;
RT "Sequence of the human insulin gene.";
RL Nature 284:26-32(1980).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=80236313; PubMed=6248962;
RA Ullrich A., Dull T.J., Gray A., Brosius J., Sures I.;
RT "Genetic variation in the human insulin gene.";
RL Science 209:612-615(1980).
RN [3]
RP SEQUENCE FROM N.A.
RX MEDLINE=80054779; PubMed=503234;
RA Bell G.I., Swain W.F., Pictet R.L., Cordell B., Goodman H.M.,
RA Rutter W.J.;
RT "Nucleotide sequence of a cDNA clone encoding human preproinsulin.";
RL Nature 282:525-527(1979).
RN [4]
RP SEQUENCE FROM N.A.
RX MEDLINE=80147417; PubMed=6927840;
RA Sures I., Goeddel D.V., Gray A., Ullrich A.;
RT "Nucleotide sequence of human preproinsulin complementary DNA.";
RL Science 208:57-59(1980).
RN [5]
RP SEQUENCE FROM N.A.
RX MEDLINE=93364428; PubMed=8358440;
RA Lucassen A.M., Bell J.I., Julier C., Lathrop M.;
RT "Susceptibility to insulin dependent diabetes mellitus maps to a 4.1
RT kb segment of DNA spanning the insulin gene and associated VNTR.";
RL Nat. Genet. 4:305-310(1993).
RN [6]

RP SEQUENCE FROM N.A.
 RC TISSUE=Pancreas;
 RX MEDLINE=22388257; PubMed=12477932;
 RA Strausberg R.L., Feingold E.A., Grouse L.H., Derge J.G.,
 RA Klausner R.D., Collins F.S., Wagner L., Shenmen C.M., Schuler G.D.,
 RA Altschul S.F., Zeeberg B., Buetow K.H., Schaefer C.F., Bhat N.K.,
 RA Hopkins R.F., Jordan H., Moore T., Max S.I., Wang J., Hsieh F.,
 RA Diatchenko L., Marusina K., Farmer A.A., Rubin G.M., Hong L.,
 RA Stapleton M., Soares M.B., Bonaldo M.F., Casavant T.L., Scheetz T.E.,
 RA Brownstein M.J., Usdin T.B., Toshiyuki S., Carninci P., Prange C.,
 RA Raha S.S., Loquellano N.A., Peters G.J., Abramson R.D., Mullahy S.J.,
 RA Bosak S.A., McEwan P.J., McKernan K.J., Malek J.A., Gunaratne P.H.,
 RA Richards S., Worley K.C., Hale S., Garcia A.M., Gay L.J., Hulyk S.W.,
 RA Villalon D.K., Muzny D.M., Sodergren E.J., Lu X., Gibbs R.A.,
 RA Fahey J., Helton E., Kettelman M., Madan A., Rodrigues S., Sanchez A.,
 RA Whiting M., Madan A., Young A.C., Shevchenko Y., Bouffard G.G.,
 RA Blakesley R.W., Touchman J.W., Green E.D., Dickson M.C.,
 RA Rodriguez A.C., Grimwood J., Schmutz J., Myers R.M.,
 RA Butterfield Y.S.N., Krzywinski M.I., Skalska U., Smailus D.E.,
 RA Schnerch A., Schein J.E., Jones S.J.M., Marra M.A.;
 RT "Generation and initial analysis of more than 15,000 full-length
 RT human and mouse cDNA sequences.";
 RL Proc. Natl. Acad. Sci. U.S.A. 99:16899-16903(2002).
 RN [7]
 RP SEQUENCE OF 1-59 FROM N.A.
 RC TISSUE=Blood;
 RA Fajardy I.I., Weill J.J., Stuckens C.C., Danze P.M.P.;
 RT "Description of a novel RFLP diallelic polymorphism (-127 BsgI C/G)
 RT within the 5' region of insulin gene.";
 RL Submitted (JUL-1998) to the EMBL/GenBank/DDBJ databases.
 RN [8]
 RP SEQUENCE OF 25-54 AND 90-110.
 RA Nicol D.S.H.W., Smith L.F.;
 RT "Amino-acid sequence of human insulin.";
 RL Nature 187:483-485(1960).
 RN [9]
 RP SEQUENCE OF 57-87.
 RX MEDLINE=71116410; PubMed=5101771;
 RA Oyer P.E., Cho S., Peterson J.D., Steiner D.F.;
 RT "Studies on human proinsulin. Isolation and amino acid sequence of
 RT the human pancreatic C-peptide.";
 RL J. Biol. Chem. 246:1375-1386(1971).
 RN [10]
 RP SEQUENCE OF 57-87.
 RX MEDLINE=71257722; PubMed=5560404;
 RA Ko A., Smyth D.G., Markussen J., Sundby F.;
 RT "The amino acid sequence of the C-peptide of human proinsulin.";
 RL Eur. J. Biochem. 20:190-199(1971).
 RN [11]
 RP SYNTHESIS.
 RX MEDLINE=75077277; PubMed=4443293;
 RA Sieber P., Kamber B., Hartmann A., Joehl A., Riniker B., Rittel W.;
 RT "Total synthesis of human insulin under directed formation of the
 RT disulfide bonds.";
 RL Helv. Chim. Acta 57:2617-2621(1974).
 RN [12]
 RP SYNTHESIS OF 57-87.

RX MEDLINE=75040007; PubMed=4803504;
 RA Naithani V.K.;
 RT "Studies on polypeptides, IV. The synthesis of C-peptide of human
 RT proinsulin.";
 RL Hoppe-Seyler's Z. Physiol. Chem. 354:659-672(1973).
 RN [13]
 RP SYNTHESIS OF 65-69 AND 70-73.
 RX MEDLINE=73161263; PubMed=4698555;
 RA Geiger R., Volk A.;
 RT "Synthesis of peptides with the properties of human proinsulin C
 RT peptides (hC peptide). 3. Synthesis of the sequences 14-17 and 9-13
 RT of human proinsulin C peptides.";
 RL Chem. Ber. 106:199-205(1973).
 RN [14]
 RP SYNTHESIS OF 84-87.
 RX MEDLINE=73161261; PubMed=4698553;
 RA Geiger R., Jaeger G., Keonig W., Treuth G.;
 RT "Synthesis of peptides with the properties of human proinsulin C
 RT peptides (hC peptide). I. Scheme for the synthesis and preparation of
 RT the sequence 28-31 of human proinsulin C peptide.";
 RL Chem. Ber. 106:188-192(1973).
 RN [15]
 RP VARIANT LOS ANGELES SER-48.
 RX MEDLINE=84016053; PubMed=6312455;
 RA Haneda M., Chan S.J., Kwok S.C.M., Rubenstein A.H., Steiner D.F.;
 RT "Studies on mutant human insulin genes: identification and sequence
 RT analysis of a gene encoding [SerB24]insulin.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:6366-6370(1983).
 RN [16]
 RP VARIANTS LOS ANGELES SER-48 AND CHICAGO LEU-49.
 RX MEDLINE=84170233; PubMed=6424111;
 RA Shoelson S., Fickova M., Haneda M., Nahum A., Musso G., Kaiser E.T.,
 RA Rubenstein A.H., Tager H.;
 RT "Identification of a mutant human insulin predicted to contain a
 RT serine-for-phenylalanine substitution.";
 RL Proc. Natl. Acad. Sci. U.S.A. 80:7390-7394(1983).
 RN [17]
 RP VARIANT PROVIDENCE ASP-34.
 RX MEDLINE=87175640; PubMed=3470784;
 RA Chan S.J., Seino S., Gruppuso P.A., Schwartz R., Steiner D.F.;
 RT "A mutation in the B chain coding region is associated with impaired
 RT proinsulin conversion in a family with hyperproinsulinemia.";
 RL Proc. Natl. Acad. Sci. U.S.A. 84:2194-2197(1987).
 RN [18]
 RP VARIANT WAKAYAMA LEU-92.
 RX MEDLINE=87058122; PubMed=3537011;
 RA Sakura H., Iwamoto Y., Sakamoto Y., Kuzuya T., Hirata H.;
 RT "Structurally abnormal insulin in a diabetic patient. Characterization
 RT of the mutant insulin A3 (Val-->Leu) isolated from the pancreas.";
 RL J. Clin. Invest. 78:1666-1672(1986).
 RN [19]
 RP VARIANT HIS-89.
 RX MEDLINE=90317021; PubMed=2196279;
 RA Barbetti F., Raben N., Kadowaki T., Cama A., Accili D., Gabbay K.H.,
 RA Merenich J.A., Taylor S.I., Roth J.;
 RT "Two unrelated patients with familial hyperproinsulinemia due to a
 RT mutation substituting histidine for arginine at position 65 in the

RT proinsulin molecule: identification of the mutation by direct
RT sequencing of genomic deoxyribonucleic acid amplified by polymerase
RT chain reaction.";
RL J. Clin. Endocrinol. Metab. 71:164-169(1990).
RN [20]
RP VARIANT HIS-89.
RX MEDLINE=85261996; PubMed=4019786;
RA Shibasaki Y., Kawakami T., Kanazawa Y., Akanuma Y., Takaku F.;
RT "Posttranslational cleavage of proinsulin is blocked by a point
RT mutation in familial hyperproinsulinemia.";
RL J. Clin. Invest. 76:378-380(1985).
RN [21]
RP VARIANT KYOTO LEU-89.
RX MEDLINE=92291307; PubMed=1601997;
RA Yano H., Kitano N., Morimoto M., Polonsky K.S., Imura H., Seino Y.;
RT "A novel point mutation in the human insulin gene giving rise to
RT hyperproinsulinemia (proinsulin Kyoto).";
RL J. Clin. Invest. 89:1902-1907(1992).
RN [22]
RP STRUCTURE BY NMR.
RX MEDLINE=91104966; PubMed=2271664;
RA Hua Q.-X., Weiss M.A.;
RT "Toward the solution structure of human insulin: sequential 2D 1H NMR
RT assignment of a des-pentapeptide analogue and comparison with crystal
RT structure.";
RL Biochemistry 29:10545-10555(1990).
RN [23]
RP STRUCTURE BY NMR.
RX MEDLINE=91242467; PubMed=2036420;
RA Hua Q.-X., Weiss M.A.;
RT "Comparative 2D NMR studies of human insulin and des-pentapeptide
RT insulin: sequential resonance assignment and implications for protein
RT dynamics and receptor recognition.";
RL Biochemistry 30:5505-5515(1991).
RN [24]
RP STRUCTURE BY NMR.
RX MEDLINE=91265527; PubMed=1646635;
RA Hua Q.-X., Weiss M.A.;
RT "Two-dimensional NMR studies of Des-(B26-B30)-insulin: sequence-
RT specific resonance assignments and effects of solvent composition.";
RL Biochim. Biophys. Acta 1078:101-110(1991).

Query Match 45.5%; Score 267; DB 1; Length 110;
Best Local Similarity 60.5%; Pred. No. 2.1e-21;
Matches 52; Conservative 0; Mismatches 0; Indels 34; Gaps 1;

```

Qy      56 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
          ||||||||||||||||||||||||||||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84

Qy      86 ----RGIVEQCCTSICSLYQLENYCN 107
          ||||||||||||||||||||
Db      85 SLQKRGIVEQCCTSICSLYQLENYCN 110

```

RESULT 7
INS_PANTR

ID INS_PANTR STANDARD; PRT; 110 AA.
 AC P30410;
 DT 01-APR-1993 (Rel. 25, Created)
 DT 01-APR-1993 (Rel. 25, Last sequence update)
 DT 15-MAR-2004 (Rel. 43, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Pan troglodytes (Chimpanzee).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Pan.
 OX NCBI_TaxID=9598;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=92219953; PubMed=1560757;
 RA Seino S., Bell G.I., Li W.;
 RT "Sequences of primate insulin genes support the hypothesis of a
 RT slower rate of molecular evolution in humans and apes than in
 RT monkeys.";
 RL Mol. Biol. Evol. 9:193-203(1992).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=22833521; PubMed=12952878;
 RA Stead J.D., Hurles M.E., Jeffreys A.J.;
 RT "Global haplotype diversity in the human insulin gene region.";
 RL Genome Res. 13:2101-2111(2003).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; X61089; CAA43403.1; -.
 DR EMBL; AY137497; AAN06933.1; -.
 DR PIR; A42179; A42179.
 DR PDB; 1EFE; 29-MAR-00.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.

FT	DISULFID	43	109	INTERCHAIN.
FT	DISULFID	95	100	
SO	SEQUENCE	110 AA;	12025 MW;	41EB8DF79837CEF5 CRC64;

```

Qy      56 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
      |||
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRREAEDLQVGQVELGGGPGAGSLQPLALEG 84
      |||
Qy      86 ----RGIVEQCCTSICSLYQLENYCN 107
      |||
Db      85 SLOKRGIVEQCCTSICSLYQLENYCN 110

```

RESULT 8

```

ID      INS_SPETR          STANDARD;          PRT;    110 AA.
AC      Q91XI3;
DT      10-OCT-2003 (Rel. 42, Created)
DT      10-OCT-2003 (Rel. 42, Last sequence update)
DT      10-OCT-2003 (Rel. 42, Last annotation update)
DE      Insulin precursor.
GN      INS.
OS      Spermophilus tridecemlineatus (Thirteen-lined ground squirrel).
OC      Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC      Mammalia; Eutheria; Rodentia; Sciurognathi; Sciuridae; Sciurinae;
OC      Spermophilus.
OX      NCBI_TaxID=43179;
RN      [1]
RP      SEQUENCE FROM N.A.
RC      TISSUE=Pancreas;
RA      Tredrea M.M., Buck M.J., Guhaniyogi J., Squire T.L., Andrews M.T.;
RT      "Regulation of PDK4 expression in a hibernating mammal.";
RL      Submitted (JUN-2001) to the EMBL/GenBank/DBJ databases.
CC      -!- FUNCTION: Insulin decreases blood glucose concentration. It
CC          increases cell permeability to monosaccharides, amino acids and
CC          fatty acids. It accelerates glycolysis, the pentose phosphate
CC          cycle, and glycogen synthesis in liver.
CC      -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC          disulfide bonds.
CC      -!- SUBCELLULAR LOCATION: Secreted.
CC      -!- SIMILARITY: Belongs to the insulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; AY038604; AAK72558.1; -.
DR      HSSP; P01308; 1LNP.
DR      InterPro; IPR004825; Ins/IGF/relax.

```

DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24 BY SIMILARITY.
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 43 109 INTERCHAIN (BY SIMILARITY).
 FT DISULFID 95 100 BY SIMILARITY.
 SQ SEQUENCE 110 AA; 12004 MW; 4511768D6622BEE5 CRC64;

Query Match 45.3%; Score 266; DB 1; Length 110;
 Best Local Similarity 57.4%; Pred. No. 2.7e-21;
 Matches 54; Conservative 1; Mismatches 3; Indels 36; Gaps 2;

Qy 50 LGTGP--RFVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
 || | |||||
 Db 17 LGPDPAQAFVNQHLCGSHLVEALYLVCGERGFFYTPKSRREVEEQGGQVELGGGPGAGL 76
 Qy 86 -----RGIVEQCCTSICSLYQLENYCN 107
 |||||
 Db 77 PQPLALEMALQKRGIVEQCCTSICSLYQLENYCN 110

RESULT 9

INS_BALBO

ID INS_BALBO STANDARD; PRT; 51 AA.
 AC P01314;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin.
 GN INS.
 OS Balaenoptera borealis (Sei whale).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Cetacea; Mysticeti;
 OC Balaenopteridae; Balaenoptera.
 OX NCBI_TaxID=9768;
 RN [1]
 RP SEQUENCE.
 RA Ishihara Y., Saito T., Ito Y., Fujino M.;
 RT "Structure of sperm- and sei-whale insulins and their breakdown by
 RT whale pepsin."
 RL Nature 181:1468-1469(1958).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR PIR; A01582; INWH1S.
 DR HSSP; P01317; 1APH.

DR InterPro; IPR004825; Ins/IGF/relax.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.
 FT NON_CONS 30 31
 FT CHAIN 31 51 INSULIN A CHAIN.
 FT DISULFID 7 37 INTERCHAIN.
 FT DISULFID 19 50 INTERCHAIN.
 FT DISULFID 36 41
 SQ SEQUENCE 51 AA; 5723 MW; 9007B50E400A7DDD CRC64;

Query Match 44.9%; Score 263.5; DB 1; Length 51;
 Best Local Similarity 92.3%; Pred. No. 2.2e-21;
 Matches 48; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSIQSLYQLENYCN 107
 ||||| |
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCASTCSLYQLENYCN 51

RESULT 10

INS_CAMDR

ID INS_CAMDR STANDARD; PRT; 51 AA.
 AC P01320;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin.
 GN INS.
 OS Camelus dromedarius (Dromedary) (Arabian camel).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Cetartiodactyla; Tylopoda; Camelidae; Camelus.
 OX NCBI_TaxID=9838;
 RN [1]
 RP SEQUENCE.
 RA Danho W.O.;
 RT "The isolation and characterization of insulin of camel (Camelus dromedarius).";
 RL J. Fac. Med. Baghdad 14:16-28(1972).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It increases cell permeability to monosaccharides, amino acids and fatty acids. It accelerates glycolysis, the pentose phosphate cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 DR PIR; A92782; INCMA.
 DR HSSP; P01317; 2INS.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism.
 FT CHAIN 1 30 INSULIN B CHAIN.

FT	NON_CONS	30	31	
FT	CHAIN	31	51	INSULIN A CHAIN.
FT	DISULFID	7	37	INTERCHAIN.
FT	DISULFID	19	50	INTERCHAIN.
FT	DISULFID	36	41	
SQ	SEQUENCE	51 AA;	5693 MW;	901E88BA085A7DDD CRC64;

```

RESULT 11
INS_CAPHI
ID INS_CAPHI STANDARD; PRT; 51 AA.
AC P01319;
DT 21-JUL-1986 (Rel. 01, Created)
DT 21-JUL-1986 (Rel. 01, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin.
GN INS.
OS Capra hircus (Goat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovoidea;
OC Bovidae; Caprinae; Capra.
OX NCBI_TaxID=9925;
RN [1]
RP SEQUENCE.
RX MEDLINE=66160119; PubMed=5949593;
RA Smith L.F.;
RT "Species variation in the amino acid sequence of insulin.";
RL Am. J. Med. 40:662-666(1966).
CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.
DR PIR; A01586; INGT.
DR HSSP; P01317; 1APH.
DR InterPro; IPR004825; Ins/IGF/relax.
DR PRINTS; PR00277; INSULINB.
DR SMART; SM00078; IIGF; 1.
DR PROSITE; PS00262; INSULIN; 1.
KW Insulin family; Hormone; Glucose metabolism.
FT CHAIN 1 30 INSULIN B CHAIN.
FT NON_CONS 30 31
FT CHAIN 31 51 INSULIN A CHAIN.
FT DISULFID 7 37 INTERCHAIN.
FT DISULFID 19 50 INTERCHAIN.
FT DISULFID 36 41

```

SQ SEQUENCE 51 AA; 5692 MW; 9007B50CDB4E7DDD CRC64;

Query Match 44.9%; Score 263.5; DB 1; Length 51;
Best Local Similarity 90.4%; Pred. No. 2.2e-21;
Matches 47; Conservative 1; Mismatches 3; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSI~~CS~~LYQLENYCN 107
| | | | | : | | | | |
Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCAGVCSLYQLENYCN 51

RESULT 12

INS_PIG

ID INS_PIG STANDARD; PRT; 108 AA.
AC P01315; Q9TSJ5;
DT 21-JUL-1986 (Rel. 01, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 10-OCT-2003 (Rel. 42, Last annotation update)
DE Insulin precursor.
GN INS.
OS Sus scrofa (Pig).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Suina; Suidae; Sus.
OX NCBI_TaxID=9823;
RN [1]
RP SEQUENCE FROM N.A.
RA Han X.G., Tuch B.E.;
RT "Complete porcine preproinsulin cDNA sequence.";
RL Submitted (MAY-1998) to the EMBL/GenBank/DDBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=Large white;
RX MEDLINE=22135958; PubMed=12140686;
RA Amarger V., Nguyen M., Laere A.S., Braunschweig M., Nezer C.,
RA Georges M., Andersson L.;
RT "Comparative sequence analysis of the INS-IGF2-H19 gene cluster in
RT pigs.";
RL Mamm. Genome 13:388-398(2002).
RN [3]
RP SEQUENCE OF 25-108.
RX MEDLINE=68286485; PubMed=5657063;
RA Chance R.E., Ellis R.M., Bromer W.W.;
RT "Porcine proinsulin: characterization and amino acid sequence.";
RL Science 161:165-167(1968).
RN [4]
RP REVISION TO 59.
RA Chance R.E.;
RL Submitted (JUL-1970) to the PIR data bank.
RN [5]
RP X-RAY CRYSTALLOGRAPHY (1.9 ANGSTROMS).
RA Blundell T.L., Dodson G.G., Hodgkin D., Mercola D.;
RT "Insulin. The structure in the crystal and its reflection in
RT chemistry and biology.";
RL Adv. Protein Chem. 26:279-402(1972).
RN [6]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RA Isaacs N.W., Agarwal R.C.;

RT "Experience with fast Fourier least squares in the refinement of the
RT crystal structure of rhombohedral 2-zinc insulin at 1.5-A
RT resolution.";
RL Acta Crystallogr. A 34:782-791(1978).
RN [7]
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
RX MEDLINE=89099318; PubMed=2905485;
RA Baker E.N., Blundell T.L., Cutfield J.F., Cutfield S.M., Dodson E.J.,
RA Dodson G.G., Crowfoot Hodgkin D.M., Hubbard R.E., Isaacs N.W.,
RA Reynolds C.D., Sakabe K., Sakabe N., Vijayan N.M.;
RT "The structure of 2Zn pig insulin crystals at 1.5-A resolution.";
RL Philos. Trans. R. Soc. Lond., B, Biol. Sci. 319:369-456(1988).
RN [8]

RP X-RAY CRYSTALLOGRAPHY (2.0 ANGSTROMS).
RX MEDLINE=92126280; PubMed=1772633;
RA Balschmidt P., Hansen F.B., Dodson E., Dodson G., Korber F.;
RT "Structure of porcine insulin cocrystallized with clupeine Z.";
RL Acta Crystallogr. B 47:975-986(1991).
RN [9]

RP X-RAY CRYSTALLOGRAPHY.
RX MEDLINE=91222450; PubMed=2025410;
RA Badger J., Harris M.R., Reynolds C.D., Evans A.C., Dodson E.J.,
RA Dodson G.G., North A.C.T.;
RT "Structure of the pig insulin dimer in the cubic crystal.";
RL Acta Crystallogr. B 47:127-136(1991).
RN [10]

RP X-RAY CRYSTALLOGRAPHY (1.65 ANGSTROMS).
RA Diao J.-S., Wan Z.-L., Chang W.-R., Liang D.-C.;
RT "Structure of monomeric porcine DesB1-B2 despentapeptide (B26-B30)
RT insulin at 1.65-A resolution.";
RL Acta Crystallogr. D 53:507-512(1997).

CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
CC increases cell permeability to monosaccharides, amino acids and
CC fatty acids. It accelerates glycolysis, the pentose phosphate
CC cycle, and glycogen synthesis in liver.
CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC disulfide bonds.
CC -!- SUBCELLULAR LOCATION: Secreted.
CC -!- SIMILARITY: Belongs to the insulin family.
CC -!- DATABASE: NAME=Protein Spotlight;
CC NOTE=Issue 9 of April 2001;
CC WWW="<http://www.expasy.org/spotlight/articles/sptlt009.html>".

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).
CC -----

DR EMBL; AF064555; AAC77920.1; ALT_INIT.
DR EMBL; AY044828; AAL69550.1; -.
DR PDB; 3INS; 09-JAN-89.
DR PDB; 4INS; 31-JUL-94.
DR PDB; 6INS; 31-JAN-94.
DR PDB; 7INS; 31-JAN-94.

DR PDB; 9INS; 15-OCT-91.
 DR PDB; 1IZA; 15-OCT-91.
 DR PDB; 1IZB; 15-OCT-91.
 DR PDB; 2TCI; 29-JAN-96.
 DR PDB; 1MPJ; 29-JAN-96.
 DR PDB; 3MTH; 29-JAN-96.
 DR PDB; 1DEI; 16-JUN-97.
 DR PDB; 1SDB; 01-APR-98.
 DR PDB; 1WAV; 28-FEB-97.
 DR PDB; 1ZEI; 16-FEB-99.
 DR PDB; 1ZNI; 28-JAN-98.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.
 DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal; 3D-structure.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 85 C PEPTIDE.
 FT CHAIN 88 108 INSULIN A CHAIN.
 FT DISULFID 31 94 INTERCHAIN.
 FT DISULFID 43 107 INTERCHAIN.
 FT DISULFID 93 98
 FT HELIX 26 46
 FT STRAND 48 48
 FT HELIX 89 94
 FT HELIX 100 106
 FT STRAND 107 107
 SQ SEQUENCE 108 AA; 11671 MW; CB4491B429858EBE CRC64;

 Query Match 44.8%; Score 263; DB 1; Length 108;
 Best Local Similarity 60.7%; Pred. No. 5.4e-21;
 Matches 51; Conservative 0; Mismatches 1; Indels 32; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
 |||
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREAENPQAGAVELGGGLGGLQALALEGPP 84

 Qy 86 --RGIVEQCCTSIQSLYQLENYCN 107
 |||
 Db 85 QKRGIVEQCCTSIQSLYQLENYCN 108

RESULT 13

INS_RABIT

ID INS_RABIT STANDARD; PRT; 110 AA.
 AC P01311;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 01-FEB-1996 (Rel. 33, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Oryctolagus cuniculus (Rabbit).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Lagomorpha; Leporidae; Oryctolagus.
 OX NCBI_TaxID=9986;

```

RN      [1]
RP      SEQUENCE FROM N.A.
RC      STRAIN=New Zealand white; TISSUE=Pancreas;
RX      MEDLINE=94179230; PubMed=8132571;
RA      Devaskar S.U., Giddings S.J., Rajakumar P.A., Carnaghi L.R.,
RA      Menon R.K., Zahm D.S.;
RT      "Insulin gene expression and insulin synthesis in mammalian neuronal
RT      cells.";
RL      J. Biol. Chem. 269:8445-8454(1994).
RN      [2]
RP      SEQUENCE OF 25-54 AND 90-110.
RX      MEDLINE=66160119; PubMed=5949593;
RA      Smith L.F.;
RT      "Species variation in the amino acid sequence of insulin.";
RL      Am. J. Med. 40:662-666(1966).
RN      [3]
RP      SEQUENCE OF 56-110 FROM N.A.
RA      Giddings S.J., Carnaghi L.R., Devaskar S.U.;
RL      Submitted (APR-1991) to the EMBL/GenBank/DDBJ databases.
CC      -!- FUNCTION: Insulin decreases blood glucose concentration. It
CC      increases cell permeability to monosaccharides, amino acids and
CC      fatty acids. It accelerates glycolysis, the pentose phosphate
CC      cycle, and glycogen synthesis in liver.
CC      -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC      disulfide bonds.
CC      -!- SUBCELLULAR LOCATION: Secreted.
CC      -!- SIMILARITY: Belongs to the insulin family.
CC      -----
CC      This SWISS-PROT entry is copyright. It is produced through a collaboration
CC      between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC      the European Bioinformatics Institute. There are no restrictions on its
CC      use by non-profit institutions as long as its content is in no way
CC      modified and this statement is not removed. Usage by and for commercial
CC      entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC      or send an email to license@isb-sib.ch).
CC      -----
DR      EMBL; U03610; AAA19033.1; -.
DR      EMBL; M61153; AAA17540.1; -.
DR      PIR; A53438; INRB.
DR      HSSP; P01308; 1TYM.
DR      InterPro; IPR004825; Ins/IGF/relax.
DR      Pfam; PF00049; Insulin; 1.
DR      PRINTS; PR00277; INSULINB.
DR      SMART; SM00078; IIGF; 1.
DR      PROSITE; PS00262; INSULIN; 1.
KW      Insulin family; Hormone; Glucose metabolism; Signal.
FT      SIGNAL          1          24
FT      CHAIN           25          54      INSULIN B CHAIN.
FT      PROPEP          57          87      C PEPTIDE.
FT      CHAIN           90         110      INSULIN A CHAIN.
FT      DISULFID        31          96      INTERCHAIN.
FT      DISULFID        43         109      INTERCHAIN.
FT      DISULFID        95         100
FT      CONFLICT        83          83      E -> Y (IN REF. 3).
SQ      SEQUENCE      110 AA;  11838 MW;  82D2975B85D77FA8 CRC64;

```

Query Match 44.8%; Score 263; DB 1; Length 110;

Best Local Similarity 59.3%; Pred. No. 5.6e-21;
Matches 51; Conservative 1; Mismatches 0; Indels 34; Gaps 1;

```
Qy      56 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
          |||||||||||||||||||||:
Db      25 FVNQHLCGSHLVEALYLVCGERGFFYTPKSRRVEELQVGQAEELGGGPGAGGLQPSALEL 84

Qy      86 ----RGIVEQCCTSIQSLYQLENYCN 107
          |||||||||||||||||
Db      85 ALQKRGIVEQCCTSIQSLYQLENYCN 110
```

RESULT 14

INS_FELCA

```
ID  INS_FELCA      STANDARD;      PRT;      51 AA.
AC  P06306;
DT  01-JAN-1988 (Rel. 06, Created)
DT  01-JAN-1988 (Rel. 06, Last sequence update)
DT  10-OCT-2003 (Rel. 42, Last annotation update)
DE  Insulin.
GN  INS.
OS  Felis silvestris catus (Cat).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Carnivora; Fissipedia; Felidae; Felis.
OX  NCBI_TaxID=9685;
RN  [1]
RP  SEQUENCE.
RX  MEDLINE=86214076; PubMed=3518635;
RA  Hallden G., Gafvelin G., Mutt V., Joernvall H.;
RT  "Characterization of cat insulin.";
RL  Arch. Biochem. Biophys. 247:20-27(1986).
CC  -!- FUNCTION: Insulin decreases blood glucose concentration. It
CC      increases cell permeability to monosaccharides, amino acids and
CC      fatty acids. It accelerates glycolysis, the pentose phosphate
CC      cycle, and glycogen synthesis in liver.
CC  -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
CC      disulfide bonds.
CC  -!- SUBCELLULAR LOCATION: Secreted.
CC  -!- SIMILARITY: Belongs to the insulin family.
DR  PIR; A01588; INCT.
DR  HSSP; P01317; 1APH.
DR  InterPro; IPR004825; Ins/IGF/relax.
DR  PRINTS; PR00277; INSULINB.
DR  SMART; SM00078; IIGF; 1.
DR  PROSITE; PS00262; INSULIN; 1.
KW  Insulin family; Hormone; Glucose metabolism.
FT  CHAIN      1      30      INSULIN B CHAIN.
FT  NON_CONS   30     31
FT  CHAIN      31     51      INSULIN A CHAIN.
FT  DISULFID    7      37      INTERCHAIN.
FT  DISULFID   19     50      INTERCHAIN.
FT  DISULFID   36     41
SQ  SEQUENCE   51 AA;  5745 MW;  9007B5096A0A7DDD CRC64;
```

Query Match 44.7%; Score 262.5; DB 1; Length 51;
Best Local Similarity 90.4%; Pred. No. 2.8e-21;
Matches 47; Conservative 2; Mismatches 2; Indels 1; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKTRGIVEQCCTSICSLYQLENYCN 107
 ||||| | : ||||| : |||
 Db 1 FVNQHLCGSHLVEALYLVCGERGFFYTPKA-GIVEQCCASVCSLYQLEHYCN 51

RESULT 15

INS_CANFA

ID INS_CANFA STANDARD; PRT; 110 AA.
 AC P01321;
 DT 21-JUL-1986 (Rel. 01, Created)
 DT 21-JUL-1986 (Rel. 01, Last sequence update)
 DT 10-OCT-2003 (Rel. 42, Last annotation update)
 DE Insulin precursor.
 GN INS.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 OX NCBI_TaxID=9615;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=83109071; PubMed=6296142;
 RA Kwok S.C.M., Chan S.J., Steiner D.F.;
 RT "Cloning and nucleotide sequence analysis of the dog insulin gene.
 RT Coded amino acid sequence of canine preproinsulin predicts an
 RT additional C-peptide fragment.";
 RL J. Biol. Chem. 258:2357-2363(1983).
 RN [2]
 RP SEQUENCE OF 25-54 AND 90-110.
 RX MEDLINE=66160119; PubMed=5949593;
 RA Smith L.F.;
 RT "Species variation in the amino acid sequence of insulin.";
 RL Am. J. Med. 40:662-666(1966).
 CC -!- FUNCTION: Insulin decreases blood glucose concentration. It
 CC increases cell permeability to monosaccharides, amino acids and
 CC fatty acids. It accelerates glycolysis, the pentose phosphate
 CC cycle, and glycogen synthesis in liver.
 CC -!- SUBUNIT: Heterodimer of a B chain and an A chain linked by two
 CC disulfide bonds.
 CC -!- SUBCELLULAR LOCATION: Secreted.
 CC -!- SIMILARITY: Belongs to the insulin family.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; V00179; CAA23475.1; -.
 DR PIR; A92413; IPDG.
 DR HSSP; P01317; IAPH.
 DR InterPro; IPR004825; Ins/IGF/relax.
 DR Pfam; PF00049; Insulin; 1.
 DR PRINTS; PR00277; INSULINB.
 DR SMART; SM00078; IIGF; 1.

DR PROSITE; PS00262; INSULIN; 1.
 KW Insulin family; Hormone; Glucose metabolism; Signal.
 FT SIGNAL 1 24
 FT CHAIN 25 54 INSULIN B CHAIN.
 FT PROPEP 57 87 C PEPTIDE.
 FT CHAIN 90 110 INSULIN A CHAIN.
 FT DISULFID 31 96 INTERCHAIN.
 FT DISULFID 43 109 INTERCHAIN.
 FT DISULFID 95 100
 SQ SEQUENCE 110 AA; 12190 MW; A574791864A4FB98 CRC64;

Query Match 44.6%; Score 262; DB 1; Length 110;
 Best Local Similarity 59.3%; Pred. No. 7.1e-21;
 Matches 51; Conservative 0; Mismatches 1; Indels 34; Gaps 1;

Qy 56 FVNQHLCGSHLVEALYLVCGERGFFYTPKT----- 85
 ||||||||||||||||||||||||||||
 Db 25 FVNQHLCGSHLVEALYLVCGERGFFYTPKARREVEDLQVRDVELAGAPGEGGLQPLALEG 84

 Qy 86 ----RGIVEQCCTSICSLYQLENYCN 107
 ||||||||||||||||||||
 Db 85 ALQKRGIVEQCCTSICSLYQLENYCN 110

Search completed: July 15, 2004, 16:36:26
 Job time : 7.38619 secs